USING GENRE TO IMPROVE WEB SEARCH

By
Mark A. Rosso

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the School of Information and Library Science.

Chapel Hill
2005

Approved by

_____
Advisor: Professor Stephanie W. Haas

_____
Reader: Professor Jane Greenberg

_____
Reader: Professor Jerry D. Saye

_____
Reader: Professor Paul Solomon

_____
Reader: Professor Elaine Toms

ABSTRACT

Mark A. Rosso: Using Genre to Improve Web Search
(Under the direction of Stephanie W. Haas)

The dissertation explores the use of genre as a document descriptor in order to improve the effectiveness of web searching. A major issue to be resolved is the identification of what document categories should be used as genres. As genre is a "folk typology", document categories must enjoy widespread recognition by their intended user groups, in order to qualify as genres. Three user studies were conducted to develop a genre palette and show that it is recognizable to users. A final study aimed to determine the palette's usefulness, in simulated web search scenarios.

The first study was a survey of user terminology for web pages. Three participants separated 100 webpage printouts into stacks according to genre, assigning names and definitions to each genre. The second study aimed to refine the resulting set of forty-eight (often conceptually and lexically similar) genre names and definitions into a smaller palette of user-preferred terminology. Ten participants classified the same 100 webpages. A set of five principles for creating a genre palette from individuals' sortings was developed, and the list of 48 was trimmed to 18 genres. The third study aimed to show that users would agree on the genres of webpages, when choosing from the genre palette. In an online experiment in which 257 participants categorized a new set of 55 pages using the18 genres, on average, over 70% agreed on the genre of each page.

The final study investigated the potential usefulness of the genre palette for web search result evaluation. Thirty-two participants performed 4 tasks. In each task, participants judged the usefulness of 20 search results & 20 webpages according to an assigned task scenario. Participants' time in judging the search results, and the stability of their judgments (as compared to their judgments of the actual pages) were compared for search results, with and without the genre of page described in each search result. The genre-annotated search results produced no significant improvement in participants' ability to make more consistent or faster relevance judgments.

Difficulties of experimental design and future directions for the work are discussed.

To my parents,

Dario Rosso (1925-1998),

Beverly Rosso (1932-1980),

and my grandmother,

Edith Bohman (1896-1978),

who always wanted the best for me…

ACKNOWLEDGMENTS

As with most endeavors of a large magnitude, one usually cannot accomplish it alone. This dissertation is certainly no exception. First, I would like to acknowledge the graciously provided assistance of those who endured being a subject in one of the first two user studies, and those who tested my websites and coded web pages for genre. In no particular order, they are: Laurie McDonald, Scott Walden, Barry Koster, Gillian Backus, Laurie Siegel, Daryn O'Shea, Amy Denham, Jane Hoppin, Thomas Pafford, Susan Sewell, Shuly Caywood, Bill Popper, Jessica Efron, Chris Rosso, Emily Rush, Joanna Burgess, Susan Barrus, Sayan Chakraborty and Hylarie Kochiras.

Also to be thanked are the 257 anonymous subjects from my online study, and the 32 of my colleagues at Meredith College who participated in the final study. I couldn't have done the statistics nearly as well or as quickly, without the help of Meredith colleague Jackie Dietz, and Chris Wiesen from the Odum Institute.

My acceptance into the doctoral program at SILS is largely due to the assistance of Jerry Saye, then Associate Dean, those long nine years ago. Thanks to Jerry's efforts, Barbara Wildemuth is also responsible for my acceptance in that she volunteered to be my initial advisor, until I could find another faculty member whose interests more closely matched my own.

My eventual advisor, Stephanie Haas, has been extremely patient with my research progress over the years, especially through the numerous crises such as my separation and

divorce, a herniated disc in my back, a broken leg, and an extended case of laryngitis. Her editing-in-progress has greatly improved the communicative quality of my work. The rest of the committee has been great as well, always being available and helpful as needed.

It was Virginia Knight at Meredith College who got me started on this long journey, first hiring me as an adjunct, and then as a full-time professor, while convincing me that I really could have a career as a college professor and earn my Ph.D. part-time. I would not have attempted it otherwise.

I am grateful to my family and friends who were understanding of all the times I had to reject offers of fun and companionship because “I have to work on my research”. Greetings from my dance friends of “Hi. How are you?” eventually became “How’s the dissertation going?” Thanks to the support of so many, I can finally now exclaim, “It’s done!”

My dad was disappointed when I dropped out of a Ph.D. program twenty-two years ago. I know that he would be proud today.

TABLE OF CONTENTS

APPENDICES Page

APPENDICES Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

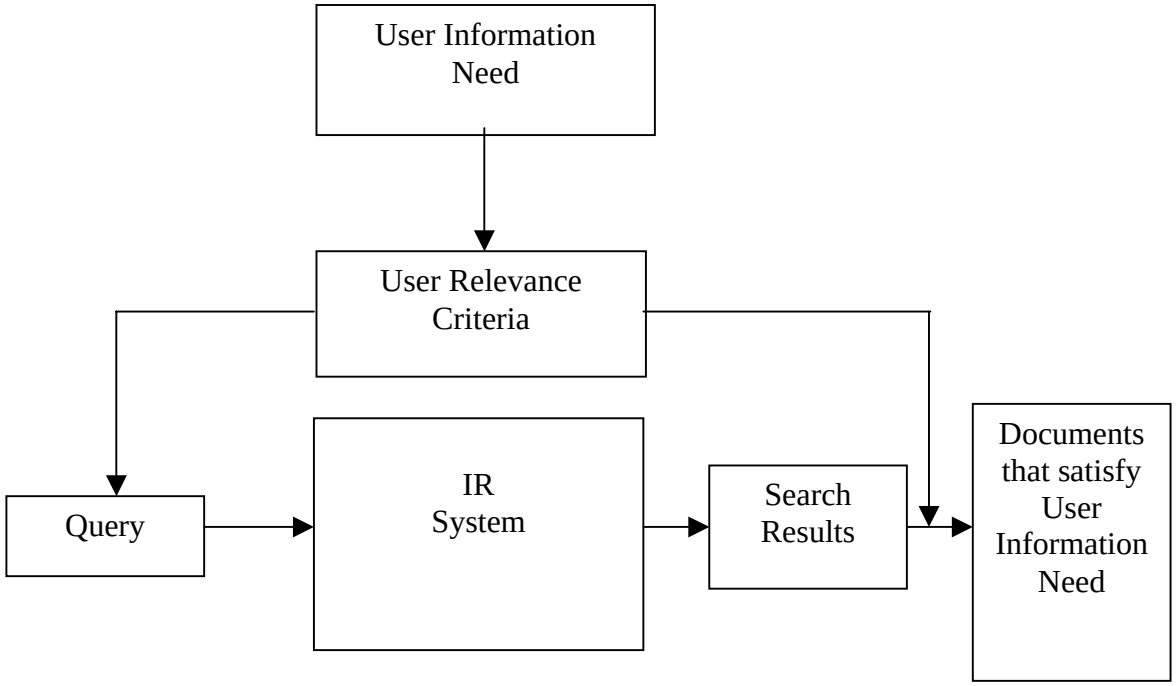
INTRODUCTION

The World Wide Web is a large, unedited repository of heterogeneous resources, with few partitions for constraining the search space. Search engines (such as Google, for example) do an admirable job of helping users find desired information, in many cases. However, in situations where relatively short, distinctive phrases (or unrelated words) that describe desired resources do not exist, or are not identified by the user, web search can be a time-consuming, futile exercise.

Furthermore, as the Web continues to grow (and assuming that search engine indices continue to attempt to keep pace), finding information can be expected to increase in difficulty. For a discussion of the relationship of collection size and search effectiveness, see for example, Blair (2002a).

In order to explain how the search process can be improved, we examine the relationship between a user's information need and the search process (see Figure 1.1).

Figure 1.1 The relationship between user information need and the search process



The Search Process and Document Descriptors

The user's information need gives rise to criteria for deciding the relevance of a document to the user's need. These criteria impact the search process at two points. First, relevance criteria shape the user's formulation of the search query, which is submitted to the information retrieval (IR) system. The typical IR system then matches the query terms to the internal representations of the documents. The representations consist primarily of the document's words, minus frequently occurring "non-content" bearing words such as prepositions, pronouns, articles, etc. The system then presents the user with a list of descriptions of documents (i.e., search results) that match the user's query. This is the second point at which user relevance criteria impact the search process. These criteria act as filters on the search results (and whatever actual documents the user chooses to inspect) for selecting documents that satisfy the user's information need. In situations in which no

documents completely satisfy the user, she may choose to continue the search by submitting a reformulated query (possibly based on a changed information need and/or changed relevance criteria, resulting from the inspection of the search results (Spink, et al. (1998), Borlund (2003a,b)).

To improve search effectiveness, it has been suggested that representing text documents primarily by their content words is inadequate (e.g., Barry (1998); Kekalainen & Jarvelin (2002a); Blair (2002a, 2002b)). When the information retrieval (IR) system matches these content words with the words in the user's query, many of the retrieved documents tend to be topically relevant to the query, but not directly relevant to the user's specific information need. For example, the query "Persian cats" may be submitted to a web search engine by someone interested in finding a Persian cat to buy. The search results returned are likely to contain pages about Persian cats that have nothing to do with selling a Persian cat. Such pages may be said to be topically relevant to the query, yet not relevant to the user's information need that underlies the query. For discussions of differing types of relevance, see for example, Borlund (2003b), Cosjin & Ingwersen (2000), Schamber (1994) and Saracevic (1975).

Numerous studies that have explored the nature of users' criteria for relevant documents (e.g., Maglaughlin & Sonnenwald (2002), Mizzaro (1998), Barry & Schamber (1998), Schamber (1994) and Schamber, et al. (1990)) have confirmed that, although topic is crucial to the relevance of a document, in most cases, non-topical criteria also play an important role in the determination of a document's relevance. In other words, to meet most user needs, topic is not enough. However, due to document representations' primary reliance on content words, IR systems do not allow the user to directly express their non-topical

criteria (other than through the standard bibliographic descriptors (typically found in document retrieval systems) such as author, title, publication date, etc.). Other examples of non-topical relevance criteria include document availability, currency, and novelty.

The identification and use of non-content (or "context") descriptors in document representations has been suggested by many (e.g., Kekalainen & Jarvelin (2002a), Blair (2002a, 2002b), Barry (1994), Cool, et al., (1993), and Schamber, et al. (1990)). Blair (2002a) describes document context as follows:

> The context of the document describes the internal or external "framework" of the document. Internal contextual information is comprised of such things as the names(s) of the author(s) of the document, its title, the type of document (memo, directive, correspondence, minutes of a meeting, etc.), date, author's affiliation, etc. External contextual information is that which cannot be gotten from the text of the document itself, such as: its place of origin, its present physical position, where it was published (if appropriate), …etc. (p. 285)

If these context descriptors correspond to, or can be related to users' non-topical relevance criteria by the users, user information needs could be more directly expressed in their queries. This could be expected to improve both the recall and precision of search results, as documents could be more accurately matched to queries. In essence, part of the users' task of filtering search results by their non-topical relevance criteria would be taken on automatically by the system (in reference to Figure 1.1).

The part of this scheme that remains to be resolved is the determination of what context descriptors to add to the document representation. The following questions should be considered:

1. How widely agreed upon are the values of a given criterion among users (or user groups)?
2. How useful is a criterion for the search tasks to be addressed by the specific IR system?

3. How easily can a criterion be identified and assigned to a document?

The first question addresses the degree to which a criterion's attribute-value is "public" or "private". (See Chapter Two for a fuller discussion of this distinction.) For example, words, as content descriptors in queries, are (more or less) effective because people generally agree on the meanings of individual words (and phrases). Even terms with multiple meanings can usually be disambiguated, based on context. Thus, it can be said that content descriptors (which can function as topical relevance criteria) are "public" knowledge: people of similar cultural backgrounds would (more or less) agree on the meanings. However, context descriptors (which can function as non-topical relevance criteria) can vary widely in the degree to which their attribute-values are considered public or private. For example, most people could agree on whether or not a document "has pictures", if given a specific document to evaluate. Thus, "has pictures" is a criterion that could be considered "public". On the other hand, the criterion of "recency" is highly situation dependent. For one individual, a document from the last ten years may be considered recent, whereas for another person, last year may not be recent enough. Thus, the judgment of a document as recent is a highly individual, or "private" event. Context descriptors' attribute values must be "public" knowledge in order to be useful as document descriptors in an IR system.

The second question addresses whether a descriptor is appropriate for the users (and their specific needs) of a given IR system. In other words, do the content descriptors correspond or relate to non-topical relevance criteria of the system's users? Will users see a relationship between their relevance criteria and these descriptors, and use these descriptors in their search queries?

The third question addresses the requirement that a viable solution must be an automatic process. There are too many web pages for manual assignment of descriptors to be

considered practical. Thus, methods for inferring document descriptors from the document contents must be developed. Fortunately, web pages have useful characteristics not found in paper documents. Web page URLs, although they consist of mostly words, may provide useful clues to non-topical relevance criteria. For example, some search engines allow searches by file type (pdf, html, jpg, etc.). Structured markup languages like HTML and XML provide information regarding document form, a non-topical attribute. Hypertext links provide relationship information between web pages, which could possibly be exploited. All of these features are currently used in typical web search engines. The innovation proposed here is the use of these features (along with document content, and other features that can be derived from document content) to infer (through an automated process) document descriptors that users can recognize as relevant to their information needs. This dissertation leaves this third question to future work.

Genre as Document Descriptor

The question asked and left unanswered at the end of the last section is basically this: what document descriptors can be incorporated into an IR system's document representation in order to improve search effectiveness? It was suggested that these descriptors be based on users' non-topical relevance criteria. This dissertation explores document genre as a viable improvement to web search engine document representation (see also (Crowston & Kwasnik, 2003; 2004) for the latest version of their earlier proposal).

Genre is a document type based on purpose, form and content. For example, a document of the "resume" genre is for soliciting employment, and is typically divided into sections with lists of descriptions of an individual's educational and work experiences, current contact information, and, optionally, references and/or hobbies. A document's genre

is typically recognized by those who create and use documents of that genre. Thus, a genre can be said to belong to a specific user group, those that share the knowledge of the purpose, form and content that is typical of that genre. For example, most young children in the U.S., and residents of countries that do not have an income tax, probably do not know what a document of the "income tax form" genre is. Various conceptualizations of genre abound in the literature of a variety of disciplines. The above description is (more or less) consistent with that which can be found in, for example, Miller (1984), Swales (1990), Yates & Orlikowski (1992) and Biber (1988; 1989; 1994).

Genre knowledge shared by those in the genre's user group usually includes typical situational characteristics. For example, one who understands the genre "income tax form" usually knows that the form is produced by the government, is filled out by someone who earns income and/or has financial losses and must report them, has a deadline for submission, requires additional schedules depending on the types of income or losses to be reported, etc.

Thus, someone who knows a particular document's genre, also knows significant things about a document, sometimes enough to a make a judgment regarding the document's relevance to an information need. For example, someone searching for advice for beginning skiers would know that a resume is not relevant, even if the document is included in an IR system's search results because it contains the phrase "beginning ski classes". Also, user studies have reported that searchers use document type in their searches of bibliographic document systems that allow it (e.g., Fidel (1991), Park (1992), Cool et al., (1993) and Tang & Solomon (1998)).

In addition to the indication of situational aspects of a document, knowledge of a document's genre can also indicate content above and beyond that indicated by the content

words in the document's representation. For example, knowing that a document is a recipe tells one that the document is about food preparation, even if the words "food" and "preparation" are not used in the recipe.

Given that genre can indicate relevance, it could possibly be used as a document descriptor in order to improve search effectiveness. Referring to Figure 1.1, genre as an indicator of relevance could impact the search process at two points. A genre recognized as a relevance indicator could be part of the user's query formulation. For example, a user could specify that only documents of that genre to be included in the search results; or, a user might decide to exclude from the search results documents of a genre deemed not to be useful. In either case, document genre is being used to constrain the search space, with the intent of improving the search results. The second point at which document description by genre could be helpful is in the search results (see Figure 1.1). Similar to the use of non-topical document descriptors noted in an earlier discussion, labeling each document description with document genre could help the user make faster and more accurate relevance judgments, possibly allowing the user to omit the viewing of some documents' full-text, which would allow the assessment of documents' relevance in a shorter amount of time. Also, genre information in the search results could be useful for query reformulation. For example, a user searching for detailed information on a medical condition, may notice a preponderance of advertisements for products in the search results, and could choose to exclude that genre from future results.

This dissertation is an exploration into the potential of genre as a document descriptor for searching the web. It is organized as follows. The next chapter is a literature review. Works in genre theory are examined and used as scaffolding for the development of a definition of web genre. Past studies of web genre are reviewed, and uses of web genre for

tasks other than searching are noted. Then, the relationship between the concepts of genre and relevance in the field of information retrieval, and its implications are discussed. Elements of cognitive theory are then used to speculate on how the use of genre may help in the information retrieval process. The final part of this literature review examines the most relevant studies of people's evaluation of document surrogates for determining the relevance of actual documents to their information needs.

Chapter Three states the research questions for this dissertation. Chapters Four through Six detail the various studies undertaken to address the research questions. Chapter Seven discusses the implications of the studies' results as a whole, and provides a summary of this work.

This page left intentionally blank.

CHAPTER TWO

LITERATURE REVIEW

Web Genre Theory

Definitions of Genre

In popular usage today, the notion of "genre" can refer to any "distinctive category of discourse of any type, spoken or written, with or without literary aspirations" (Swales, 1990, p. 33). Examples include the theatrical musical, the mystery novel, the documentary film, country western music, the cooking recipe, and even the typical exchange for buying meat at the butcher shop. Academic literature of genre theory spans a wide range of time periods and disciplines. But, essentially, genre is a classification of discourse. Disciplines concerned with genres have varying purposes for their classifications (Bhatia, 1993). For example, sociologists use genre to explore social roles, group purposes, professional and organizational preferences, and cultural constraints (Bhatia, 1993). Psychologists construct text classifications based on the mode of cognitive processing required by different types of texts (Fayol, 1991). The influence of varying text structures of genres on reading comprehension is one area of study for cognitive psychologists. Linguists have used genre to associate specific features of language with certain types of writing (Swales, 1990; Bhatia, 1993). Genre, in the area of rhetorical study, is used to teach the social and cultural factors of language in use (Freedman & Medway, 1994). More recently, genres in information systems

have been studied in order to understand system usability in the field of technical communications research (e.g., Spinuzzi, 1999a, 2000). Williams (2004) studied how genre can be used to affect change in an organization's software engineering process.

I want to make a distinction between different senses in which I use the term "genre". Sometimes it refers to genre theory as a whole, e.g., as it is used in the previous paragraph. The term can also refer to specific classifications such as the mystery novel genre. A specific pre-defined set of genres will be referred to simply as a "genre palette" (a phrase attributable to Karlgren, et al., 1998). I prefer it to Orlikowski and Yates' (1994) term, "genre repertoire" which implies a more complete set of genres for a social community. Other frameworks for describing assemblages of genres include genre systems, genre sets, and genre ecologies (Spinuzzi, 2004).

The concept of genre has evolved over time in several noticeable ways. It has broadened from fixed classifications of literary works or specific rhetorical situations to open, flexible systems in which just about all types of discourse can be placed. Types of discourse include written and spoken communications, interaction (as in personal conversations) and monologue, and prepared and spontaneously delivered communications. Genres may have broadly agreed upon names such as a "comic book", or be un-named (but still recognizable) like an explanation of the floor plan of a residential apartment. In this chapter, I present a review of genre theory from the literature of rhetoric, linguistics, and organizational communication, followed by a discussion of the relationship to the use of genre in improving retrieval on the web.

Rhetorical Genre. The Early Greeks, most notably, Aristotle, used categories to classify both dramatic works (poetics) and types of public discourse (rhetoric) (van Dijk,

1985). The intention was to give names to these works so that their production and criticism could be taught to others (Kennedy, 1991). More recently, in her widely cited work, Miller (1984) noted that in the area of rhetorical theory, a profusion of genre definitions have been proposed. These genres have been classified by similarities in strategies, forms, audience, mode of thinking, and rhetorical situations. Miller contended that genres should contribute to an understanding of how discourse works to "help [us] account for the way we encounter, interpret, react to, and create particular texts." (p. 23) She characterized genre as social action.

> Genre refers to a conventional category of discourse based in large-scale typification of rhetorical action; as action, it acquires meaning from situation and from the social context in which that situation arose. (p. 37)

Essentially, Miller's classification favors pragmatics (context) over syntactics (form) or semantics (substance). Miller considered this emphasis as best reflecting rhetorical practice. However, form and substance are still important, as Miller considered genre to be a fusion of situational, formal, and substantive features.

Miller defined situation not as objective (existing in physical reality) or subjective (existing separately in the mind of each individual), but as a social construct, a type. This typification is a recognition of recurring rhetorical situations in society. Thus, genre is a situation-type, which includes typification of participants. Miller explained that:

> "Successful communication would require that the participants share common types; this is possible in so far as types are socially created (or biologically innate). (p. 29)

For Miller, learning genres means to "learn to understand better the situations in which we find ourselves and the potential for failure and success in acting together" (p. 39). She offered eulogizing, apologizing, and recommending one person to another as examples of

genres. In summary, Miller's is a scheme based on context, which can accommodate new genres as they emerge over time.

Genre from a Linguistic Perspective. Swales (1990), an applied linguist focusing on the use of English in academic and research settings, defined genre as comprising "a class of communicative events, the members of which share some set of communicative purposes." (p. 58) Like Miller, Swales emphasized pragmatics over form and substance.

> Communicative purpose is both a privileged criterion and one that operates to keep the scope of a genre as here conceived narrowly *focused on comparable rhetorical action*. (p. 58, italics added)

More recently, Askehave and Swales (2001) have noted that the use of communicative purpose as a criterion for genre identification can be problematic in some types of situations, e.g., document types with multiple purposes, or types whose purposes that may not be clear to the analyst. The rationale for communicative purpose as criterion was slightly modified:

> We thus suggest that purpose (more exactly sets of communicative purposes) retain the status as a 'privileged' criterion, but in a sense different to the one proposed by Swales. It is no longer privileged by centrality, prominence or self-evident clarity, nor indeed by the repeated beliefs of users about genres, but by its status as reward or payoff for investigators as they approximate completing the hermeneutic circle (p. 210, Askehave & Swales, 2001).

Swales (1990) also explained how form and substance are associated with genre types through content schemata and formal schemata. These schemata are derived from "our assimilated direct experience of life and … our assimilated verbal experiences and encounters." (p. 83). So, like Miller, Swales' genres are based primarily on purpose/action, and also involve unique combinations of form and substance.

However, Swales' definition becomes more restrictive as he stated that the genres' communicative purposes are "recognized by the expert members of the discourse

community", a group defined by a set of common goals, mechanisms for communication among and participation by members, and a "reasonable ratio between novices and experts." (p. 24-27) Miller's genres also involved shared knowledge of situation-types between participants but there were no requirements that the participants (e.g., the producer and the reader of a genre instance) interact with each other, or even know each other. In Swales view, genres "belong" to a discourse community. For example, in a writing class for graduate students from diverse disciplines, Swales taught "Academic Correspondence" genres such as memos to dissertation committee members, and requests to academics working elsewhere. Clearly, each has a communicative purpose, that is recognized by an expert member of the discourse community (in this case, the academic community of each student's specific discipline). Academic disciplines have common goals (e.g., teaching, research), mechanisms for communication and participation (e.g., journals, conferences) and a mix of novices and experts (from graduate students to senior professors).

Like Miller, Swales shares a distaste for fixed (permanent) a priori classifications, noting their limited use for archival or typological convenience, rather than as a tool for genre discovery. However, Swales did make a case for the value of a priori classifications as an intermediate step in developing a posteriori classifications that are informed by observation and interaction with the specific discourse community. Finally, Swales elaborated on the cultural aspect of genre, observing that genres "vary in the extent to which they are likely to exhibit universal or language-specific tendencies" (p. 64, see also Bhatia (1993)).

Bhatia (1993), building on the work of Swales, proposed a similar conceptualization of genre, but with a focus on professional and academic writing. He indicated that this definition is intended strictly for non-fiction genres.

> …it [genre] is a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs. (p. 13)

However, Bhatia did not give any special status to form and substance. They were mentioned as just a few of the many factors characterizing genre.

> Although there are a number of other factors, like content, form, intended audience, medium or channel, …it [genre] is primarily characterized by the communicative purpose(s) that it is intended to fill. (p. 13)

Also, Swales' (1990) concept of "discourse community" is referred to by Bhatia as "professional or academic community" and "specialist community," but no strict requirements for the community such as individual participation or interaction were given. Bhatia's overall treatment is in the spirit of a "how-to" manual, as his emphasis on the analyst's need for "specialist information" from a "practicing member of the disciplinary culture in which the genre is routinely used" exemplifies. (p. 34)

As an example genre, Bhatia offered the sales promotion letter: "an unsolicited letter addressed to a selected group of prospective customers…in order to persuade them to buy a product or service." (p. 45). Bhatia emphasized the tactical cognitive structure of genres, identifying the seven characteristic "moves" of the sales promotion letter: establishing credentials, introducing the offer, offering incentives, enclosing documents, soliciting response, using pressure tactics, and ending politely.

Ferguson's (1994) perspective on genre is part of a larger interest in the study of language variation as a means to understanding "conventionalization in language", "the

process by which members of a community somehow come to share the sound-meaning pairings that constitute their means of verbal communication, in spite of the fact that no two speakers speak exactly the same way and the shared language keeps changing." (p.15)

He offered a definition of genre that combines the syntactic, semantic, and pragmatic aspects in a different way than those we have reviewed so far:

> A message type that recurs regularly in a community (in terms of semantic content, participants, occasions of use, and so on) will tend over time to develop an identifying internal structure, differentiated from other message types in the repertoire of the community. (p. 21)

The primary emphasis is on action (the message type). However, the characteristic of form (an identifying internal structure) appears to have more weight than the others mentioned. As part of his exposition, Ferguson reviewed a sociolinguistic textbook, which uses the term "message-form" interchangeably with genre. Unfortunately, there is no elaboration on the concept of community. And, finally, Ferguson stated that "genres may pass relatively easily from one speech community to another." Thus, unlike Swales' genres, Ferguson's do not belong to a community. In fact, he added that some genres are even shared across languages.

Biber (1989), also from a linguist's perspective, referred to genres as a "folk typology", and noted that they are "assigned on the basis of use" (Biber, 1988, p. 170). Genres are text categories, defined and distinguished on the basis of systematic external criteria (Biber, 1988). These external criteria consist of "subject-matter, purpose, rhetorical structure, and style in addition to situational parameters". These situational parameters include the relation between the communicative participants (of which shared background knowledge is a factor), the relation of the participants to the "external context", and other factors (Biber, 1988; 1994). Biber's conception of genre appears to be more general than that of Swales and Bhatia, as Biber's genres are "readily distinguished by mature speakers of a

language," a less restrictive definition of a genre's discourse community (Biber, 1989). Overall, Biber's concept of genre is consistent with the other reviewed work. It is based primarily on pragmatics ("use"), associated with a recurring situation ("systematically based") with characteristic substance ("subject-matter") and form ("rhetorical structure and style") – Biber uses the term "form" specifically as a below-the-text-level descriptor.

Schryer (2002) described the current state of genre theory as a " 'text in context' approach to examining recurring linguistic events" (p. 80). However, she distinguished between the rhetorical approaches and the linguistic approaches. The rhetoriticians have emphasized the "context" in "text in context", while the linguists have put more emphasis on the "text". Schryer advocates a more balanced emphasis:

> We need genre research that provides both participant accounts as well as analytical, close readings of texts that instantiate a genre. Based on such accounts, I believe that we will be able to more closely document the resources available to a genre and interrogate the way agents strategically use genres and their resources in specific contexts. Consequently, we will be able to see more clearly the relationship between genres and issues of power. (p. 74).

Genre as Organizational Communication. A widely cited work by researchers of digital genres is (Yates & Orlikowski, 1992). The authors essentially extended Miller's (1984) work to the study of genres in organizations. A genre of organizational communication was defined as "a typified communicative action invoked in response to a recurrent situation." They defined "recurrent situation" in terms of the organization: "the history and nature of established practices, social relations, and communication media."

The authors explained that a communicative action is identified as an instance of a genre "within the relevant social community." They used the concept of "normative scope" to define the relevant social community. Normative scope refers to the extent to which the

social norms of a recurrent communicative situation (along with its characteristic substance and form) are shared. Examples of normative scope varying from broad to specific are offered: widely accepted in most advanced industrial countries, specific to organizations in particular cultures, specific to certain industries and occupations, distinct organizational or corporate cultures, and intra-organizational groups such as departments and teams. Thus, the relevant social community of a genre is comprised of those who share the social norms of the recurrent situation.

The authors noted that a genre at a lower level of abstraction can be considered to be a subgenre of a higher-level genre. For example, the genre of "presidential inaugural addresses" is a subgenre of the "public speech" genre. The authors related the concept of normative scope to a genre's level of abstraction. They posited that genres with a broad normative scope are most likely to be at a high-level of abstraction, and vice-versa. Thus, generally speaking, broader genres will be recognized by a larger community than narrower genres.

Spinuzzi (1999a) combined genre theory and activity theory to produce a theory of artifacts for use in the design and evaluation of information systems (IS). His notion of genre differs from the other literature reviewed here, in two main respects. First, in his theory of artifacts, genre may refer to both textual and non-textual types of "artifacts" (e.g., mouse cursors, checkboxes, and coffee cups, Spinuzzi (1999a)). Genre instances are seen as tools used in an activity to accomplish something. They are not always communications between people. Second, genres can thus be a private phenomenon for an individual "(such as aides memoire or scribbled notes)" (Spinuzzi, 2004). These differences may make sense in the context of IS design and evaluation, but are generally inconsistent with the other contexts of

genre theory use reported in this review. Typically, genres have a communicative purpose (i.e., do not refer to coffee cups), and are socially recognized.

Spinuzzi's dissertation (1999a), which fully explains the theory and its context, defines genres as "relatively stable, historically developed, socially distributed collections of rules" (p.45).

> Collections of rules thus provide a sort of social memory, a way for subjects within activity networks to share and stabilize the strategies they have developed for dealing with artifacts. Genre provides a unit of analysis for the coherent collections of rules surrounding an artifact. (p. 45)

A later sentence implies a more familiar view of genre implying recurring situations and shared knowledge:

> A genre provides a relatively stable, historically developed set of habits that is easy for authors/developers to produce and readers/users to interpret. (p. 45)

However, four subsequent papers are generally less explicit about the theory's context, and offer varying definitions for genre: "typified forms that people have developed to communicate within an activity" (Spinuzzi, 1999b), "artifact types and the interpretive habits that have developed around them" (Spinuzzi & Zachry, 2000), simply "artifact types" (Spinuzzi, 2002), and finally

> *Genre* is a way of talking about how people regularly interpret and use texts. I mean "texts" broadly speaking: we talk about genres of literature, music, architecture, speech, and even computer interfaces. Computer interfaces and related technological artifacts can be and have been productively examined in terms of genre (p.110, Spinuzzi, 2004).

In summary, this unique application of genre theory, together with various treatments across several papers, may serve to further confuse the notion of genre as it appears in the literature. The dissertation included a good review of Bakhtin's speech genre theory (Spinuzzi, 1999a).

In another recent series of papers, Williams (2003a, 2003b, 2004) applies genre theory to a common work situation (which she terms “genre dumping”) in which one or more interacting workgroups are instructed to read and write documents of genres of which they are unfamiliar. The cited instance was one in which a software development group with experience in traditional software engineering techniques was learning object-oriented systems development, and needed to read and produce “use case” documents, rather than the types of development documents with which they were familiar. Williams questioned the adequacy of genre theory to explain these types of situations. Further, she questioned whether recurrent, typified action should be a criterion to be used in genre identification.

Genre taxonomy. Several researchers have called for a taxonomy of discourse types in order to understand different kinds of language and their inter-relationships, that would allow a “systematic study of verbal repertoire” and comparison of discourse types (Biber, 1994). Biber noted that “one important use of a taxonomic framework is to specify the level of generality.” This would facilitate comparison of discourse types, or indicate where comparison is inappropriate. Biber proposed a comprehensive framework “to provide a precise specification of both the level of generality and the particular situational characteristics.” Biber also provided an excellent review of previous taxonomic frameworks proposed by various linguists. He summarized it as follows:

> Previous frameworks for the situational characterization of registers have had one of two primary goals: classification or description. Classificatory frameworks are based on a closed set of discrete distinctions, so that the register category of any text can be specified. These frameworks typically include only a few general parameters and distinguish among only three or four major text categories. In contrast, descriptive frameworks attempt to provide complete situational characterizations of texts and registers; to accomplish this, they have utilized many open-ended parameters and thus

have not been suitable for classificatory purposes. (Biber, 1994, p. 37) [1]

His framework is essentially a faceted classification (Kwasnik, 1999), containing twenty-four facets in all (see Table 2.1).

[1] Due to lack of agreement in the linguistics literature regarding the definitions and relationships between terms that indicate situational language variation, Biber used the word "register" to cover for all such terms (including genre). (For a complete discussion regarding this terminology situation, see (Biber, 1994, p. 51-53)).

Table 2.1. Situational Parameters of Register Variation (reprinted from (Biber, 1994))

(parameters with asterisks may contain open-ended sets of values)

I. Communicative Characteristics of Participants Addressor(s):
- A. Addressor(s):
  Single/plural/institutional
- B. Addressee(s):
  1. Self/Other
  2. Single/plural/unenumerated
- C. Audience: yes/no

II. Relations Between Addressor and Addressee
- A. Social role relations-relative status and power of addressor and addressee
  Addressor has more power/equal status/addressee has more power
- B. Extent of shared knowledge
  1. Specialist knowledge of topic: high/low
  2. Specific personal knowledge: high/low
- C. Interactiveness: extensive/slight/none
- D. Personal relationship: like, respect, fear: kin, friends, enemies, colleagues, etc.

III. Setting
- A. Characteristics of the place of communication
  1. Private/public
  2. Domain:
     Business and workplace
     Education and academic
     Government and legal
     Religious
     Art and entertainment
     Domestic and personal
     Other
  3. Audio/visual mass media (television, radio, cinema)
- B. Extent to which place is hared by participants:
  immediate/familiar/removed
- C. Extent to which time is shared by participants:
  immediate/familiar/removed
- *D. Specific place and time of communication

IV. Channel
- A. Mode (primary channel):
  written/spoken/signed/mixed/(other)
- B. Permanence:
  recorded/transient
- C. Medium of transmission:
  If recorded:
  1. Taped/transcribed/typed/printed/handwritten/e-mail/other
  2. Published/unpublished
  If transient:
  3. Face-to-face/telephone/radio/TV/other
- D. Embedded in a larger text from a different register: yes/no

(Table 2.1 continued on next page)

Table 2.1. Situational Parameters of Register Variation (continued)

V. Relation to Participants to the Text
- A. Addressor-production circumstances: revised or edited/scripted/planned/on-line
- B. Addressee-comprehension circumstances: on-line/self-imposed time constraints
- C. Addressor's and addressee's personal evaluation of text: important, valuable, required, beautiful, popular, etc.
- D. Addressor's attitudinal stance toward the text:
  1. Emotionally involved/removed
  2. Reverence/everyday
  3. Excitement

  etc.
- E. Addressor's epistemological stance toward the text: belief, conviction, doubt, etc.

VI. Purposes, Intents, and Goals
- A. Factuality: (Purported to be) based on fact/speculative/imaginative/mixed
- B. Purposes:
  1. Persuade or sell: high/medium/low
  2. Transfer information: high/medium/low
  3. Entertain/edify: high/medium/low
  4. Reveal self (including expression of personal feelings, attitudes, or efforts at enhancing interpersonal relations): high/medium/low

VII. Topic/Subject
- A. Level of discussion: Specialized/general/popular
- *B. Specific subject: finance, science, religion, politics, sports, law, people, daily activities, etc.

The facets are not ranked in order of importance. One important feature of the framework is that, by leaving one or more of the facets unspecified, one can specify genres at differing levels of abstraction.

> For example, "writing" is a register at an extremely high level of generality in that only one parameter[2] is specified: primary channel. "Planned discourse" is a register at the same level of generality, but it differs from "writing" in that the primary channel parameter is unspecified, while the only specified parameter is for production circumstances. (p.42)

[2] Biber used the term, parameter, for facet.

I would like to elaborate on two of the facets. "Extent of shared knowledge" includes the previously discussed notion of normative scope (Yates & Orlikowski, 1992). Biber explained that:

> Shared knowledge can refer to specialist knowledge of particular topics or to specific personal background knowledge. (It can also refer to cultural world knowledge, which would be relevant in cross-cultural communication.) (Biber, 1994, p. 42)

Another facet indicates whether the text in question is embedded in another text of a different register. Miller (1984) also alluded to this type of situation in stating that genres can potentially be ordered hierarchically in two distinct ways: as varying levels of abstraction or as part/whole relationships. For example, a book and a novel are examples of varying levels of abstraction. The latter is a sub-type of the former, in that all novels are books, but not all books are novels. However, a novel could have a job interview (another genre) embedded in it, exemplifying a part/whole relationship.

Defining Web Genre

"Genre" is, admittedly, a "squishy" term, as are many of the terms used in the genre definitions: purpose, situation, use, substance, form, external criteria, etc. Despite the lack of terminological precision, the purpose of this discussion is to illustrate where and how the application of genre theory can contribute to web-based information retrieval: in the augmentation of the user interface, and in the document representations for automated genre prediction of web documents. We start by looking at some of the common threads in the various conceptualizations of genre.

Although the selected works agree that genres have characteristic form and substance, overwhelmingly, all the genre definitions have been based on pragmatics

(use/purpose/action). Biber (1989) named these “external criteria” which are readily distinguished by “mature speakers of a language.”

Swales (1990) and Bhatia (1993) offered more restrictive definitions than mature speakers (i.e., the discourse community and the specialist community, respectively). This difference in restrictiveness may be explained by Yates & Orlikowski’s (1992) concept of normative scope. Swales’ and Bhatia’s purposes for using genres primarily involved the study and teaching of English for specific purposes, not general English. Biber’s genres are, most likely, conceived at higher levels of at higher levels of abstraction, in which the genre knowledge is shared by a wider group of people. Thus, by noting the principle of normative scope, there is no inconsistency between the various conceptualizations of a genre’s communicative participants.

In terms of web retrieval, what is important about genre is that it is a typified something, and that the users have knowledge about what a specific genre looks like, and what it means. Whether this “typified something” is action, purpose, use, situation, or something else, is probably immaterial. If users know what a genre instance looks like, they can recognize it as belonging to the genre. If they understand the genre (and its typical use, situation, etc.), it can help them to understand the retrieved web page, and to help them judge its relevance to their tasks. The people who have this knowledge in common, this “shared genre knowledge,” are essential to the definition of a genre. Without it, a genre goes unrecognized. It cannot be created or used (as genre).

Swales (1990) named this group of people with shared genre knowledge as the genre’s “discourse community”. However, I contend that, for our purposes, Swales’ defining characteristics of a discourse community are unnecessary for a group of people to have

shared genre knowledge. In terms that Miller (1994) attributes to Harre (1981), the group does not need to be "relational", but merely "taxonomic".

> Members of taxonomic collectives have similarities, perhaps even shared qualities or beliefs, but these are shared only in the sense of being common to the members, who have no real interrelations with each other. The collectivity exits in the mind of the classifier. Members of relational collectives, in contrast, have real relations with each other, by means of which active sharing occurs, and the collective itself has a structure: it is differentiated. (Miller, 1994, p. 73)

In lieu of the defining characteristics, I would define the group simply as those who have the shared genre knowledge. And because those who use the web are popularly called "web users", this group could be referred to as the genre's "user group". Thus, each genre is partially defined by its user group, those that are familiar with the gist of that genre.

How does this concept of "user group" apply to the web? Is the web one large user group, or is it a collection of many user groups? I propose that the answer to both questions is "yes". The knowledge required to understand some genres could be shared by (virtually) all web users (of some basic experience level). For example, because existing genres can migrate from traditional media to the web (e.g., Shepherd & Watters, 2004), web users familiar with traditional media could recognize them. However, other genres may require knowledge that is only shared by subsets of web users. Using the terminology of Yates and Orlikowski (1992), genres on the web may have differing "normative scopes". In summary, a genre on the web is a pragmatic type (with corresponding form and substance), recognized by the genre's user group. Next, we explore web page classifications in the research literature and determine whether they meet our definition of web genres.

Web Genres

Web Page Classification Schemes

Many web page classification schemes have been proposed. In this chapter, the classifications are evaluated with respect to the definition of web genre (defined previously) and the classifications' potential usefulness to document retrieval is discussed. Important characteristics of web genre palettes are highlighted. Finally, studies of user recognition of web genre, and uses of web genre other than for document retrieval, are reviewed.

Motivated by the internet search problem (Chen, et al., 1996), Haas and Grams (1998a) created a classification system for web pages and link types. The web page categories they created bear some similarities to the resource type descriptors in the (then current) Dublin Core proposal (Dublin Core Resource Types, 1998). Haas and Grams also gave some evidence that the categories might be recognizable to users. In a random sample of 75 web pages (out of 331 pages collected from a major search engine's random feature), inter-coder agreement of page types was 88%. It is also possible that Haas and Grams' link type classifications might be useful in determining page type detection. No search experiments to validate the usefulness of the classifications in searching were undertaken.

Like Haas & Grams (1998a), Crowston & Williams (2000) performed a content analysis on a sample of 837 web pages (randomly chosen from a larger sample of 8000, provided by a major search engine). However, their set of genres was based on general knowledge of genres, using definitions from the Oxford English Dictionary, and part of the hierarchical scheme from The Art and Architecture Thesaurus (Petersen, 1994). Crowston & Williams were studying genres on the web from the standpoint of social phenomena. Specifically, they investigated the assertions by Orlikowski & Yates (1994) that the

introduction of a new communications medium, like the Web, would result in genres that were reproduced or adapted from genres in existing media.

Inter-coder agreement on the double-coded portion of Crowston and Williams' sample ranged from 68-78%, depending on agreement criteria. There was no investigation into the genres' viability for improving search. However, the fact that the genres were based on an existing scheme, and level of inter-coder agreement suggest that the genres may be recognizable to some web users.

A brief, but comprehensive, proposal by Kwasnik and associates (Kwasnik, et al., 2000) outlined a strategy for applying genres to web retrieval. After classifying a pseudo-random sample of web pages, they proposed a descriptive study of web users' actual search behaviors. The goal of the study would be to identify "user-based genres, content-based relevance indicators of the genres, and the language users employ to label the genres" (p. 25). Then, a faceted classification (Kwasnik, 1999) of web genres would be constructed, and automatic genre classification methods using both heuristic and machine learning techniques would be developed. Finally, they would conduct research into how genres can be used in the user interface, and comparison testing of web search with and without genres. Expanded detail and rationale for the proposal is provided in (Crowston & Kwasnik, 2003; 2004).

In a preliminary study based on Kwasnik and associates' proposal, Roussinov, et al. (2001) used the a priori genres identified by Crowston and Williams (2000), and supplemented them with genres developed from users' genre assessments of 1,234 webpages collected during actual searches. Data was collected from 184 subjects, approached and interviewed "in campus computer labs, public libraries, and workplace settings engaged in searching the web." The majority worked or studied in an academic environment, and rated

their Web-use proficiency as above average. The results of the study included a list of web search purposes and associated web genres. For example, genres associated with the purpose of scholarly research included tables of contents, articles, topical home pages, essays, etc. Roussinov, et al. (2001) felt that there was "general agreement among web genres perceived by different people" and that "the ability to search for documents by genre could be useful." They reported that the implementation of automated recognition techniques is underway.

Nilan and colleagues (Nilan, et al., 2001) conducted a study similar to the above with one major difference: the genre palette was to be derived entirely from a content analysis of users' genre assessments of webpages collected during actual searches. There would be no a priori genres. They used methods similar to those used by Roussinov, et al. (2001) with the exception that additional user responses to an online version of the survey were solicited from mailing lists and Usenet newsgroups on topics "within the scope of Web development, usability, search engines, and online searching." Data from 1,351 webpages assessed by 250 users were collected. Unfortunately, they found the users' terminology too vague to be useful. For example, the term "list" was used "to refer to a number of structurally and functionally different pages" (p. 336). The authors concluded that "without reliable patterns in the usage of terms, we cannot reliably classify documents such that they can be accessed through the same terms." In light of this (and other methodological problems), analysis of the data set was abandoned.

In another work specifically aimed at the web search problem, Karlgren and associates (Karlgren, et al., 1998; Dewe, et al., 1998) developed DropJaw, a fully-functioning search-interface prototype that makes use of web genres. They solicited input from 648 students, researchers and teachers at two Swedish universities, asking what web genres the

respondents were aware of. A preliminary palette was constructed based on the users' input. Genres included personal home pages, commercial home pages, searchable indices, journalistic materials, reports, other running text, FAQs, link collections, other listings and tables, asynchronous multi-party correspondence, and error messages. The researchers were not able to construct algorithms that could identify page genres automatically with an acceptable level of accuracy. However, the researchers were optimistic that, with improvements to the algorithms and a better-defined genre palette, results would improve.

Like Karlgren and associates, Stein and Meyer zu Eissen (2004) developed a genre palette based on user input. Again, user input was obtained from a questionnaire, this one given to 286 students at the authors' university. There was no user validation study of the resulting, eight broad genres: help, article, discussion, shop, portrayal (non-private), portrayal (private), link collection, and download. The authors also experimented with several algorithms to automatically classify an 800-page test corpus into genres. No details were given as to the source of the corpus or how the pages' "gold standard" genres were assigned. The authors claim accuracy of "about 70%". Unlike Karlgren, et al., the categorized pages were not used as the basis of a web searching experiment.

Another European study reported the development of a 1.3 million-page corpus of German language web pages from German universities (Rehm, 2002). The eventual size of the corpus was estimated to be over three million pages. Building on the work of Haas & Grams (1998a, 2000), Rehm (2002) proposed that an instance of a web genre be composed of a set of compulsory and optional "web modules". For example, for a page to be part of the "Academic's Personal Homepage" genre, it must have: an indication of school affiliation; the page owner's name; contact information of which an email address is compulsory but postal

address, phone number, secretary's phone number, fax number, room number, and office hours are optional; and a list of publications. Optional elements include a link to a version in a different language, a photo of the owner, etc. Thus, searching could be performed both by genre, and by genre module within genre.

The study presented a tentative genre palette developed from a content analysis of a 200-page random sample of the larger corpus. With 35 genres and sub-genres, no mention was made of inter-code agreement. There was also no investigation into the user recognition of the genres or their viability for improving search.

Lee and Myaeng (2002) experimented with a corpus of seven, broad genres, which seem to have been developed on an ad hoc basis (as opposed to the use of user input, as in, e.g., Dewe, et al., 1998; Karlgren, et al., (1998)). Genres included editorial, reportage, review, research paper, homepage, FAQ, and product specification. Complete inter-coder agreement for the genres was implied: "each document was examined by at least two people for inclusion in the collection as well as in the designated genre…classes" (p. 147). Their best classifier achieved per genre precision ranging from 77-98%, and per genre recall ranging from 73-98%.

Björneborn (2004) created a genre classification based on 530 academic web pages "in order to enable the later identification of what types of web pages provide transversal links in an academic web space..." (p. 152). The pages are from 10 "paths" in "strongly connected components" of a 2001 crawl of 109 academic websites in the UK. Using an interative process, the author created a palette of 17 broad groups of genres that he termed "meta genres", complete with definitions, examples and a prioritized categorization order. Neighboring pages were consulted when a page's context was unclear. The meta genres were

divided into two groups: 9 institutional meta genres and 8 personal meta genres. No measure of inter-coder reliability was attempted.

Characteristics of Web Genre Palettes

These studies, as a whole, demonstrate important ways in which genre palettes may differ from one another. Levels of abstraction, user groups, mutual exclusivity, exhaustivity, the distinction between page types and genres, and the unit of genre analysis will be discussed.

As mentioned previously (Yates & Orlikowski, 1992), genre palettes may be specified at varying levels of abstraction. Haas & Grams (1998a) observed that inter-coder agreement was lower for page types at lower levels of abstraction. Two other studies (Crowston & Williams, 2000; Roussinov, et al., 2001) recognized that many inter-coder disagreements were of a hierarchical nature: for example, the coded genres were two sub-genres of the same genre, or one coded genre was a sub-genre of the other coded genre. Thus, it seems reasonable to expect that the addition of genres at a lower level of abstraction to a genre palette could make it harder to verify that the palette is recognizable to users, depending, of course, on the specific user group. How the lower levels of abstraction would alter the effectiveness of the palette for information retrieval may well depend on characteristics of the user group(s), the tasks, the size and nature of the collection, and even the specific genres themselves.

Most of the studies reviewed here propose only broad genres, and do not explicitly identify the genres' user groups, a critical part of a genre's definition. For example, both Crowston & Williams, (2000) and Roussinov, et al., (2001) characterized their web genres as "the result of interactions between communities". They further stated that their "definition of

genre relies on social acceptance", but they do not say by whom they should be accepted. Karlgren and associates (Dewe, et al., 1998; Karlgren, et al., 1998), made a similar omission in their definition of genre: "a set of documents with a perceived consistent tendency to make the same stylistic choices…or if it has an established communicative function, [it is called a] functional style". They did not define by whom the consistency is perceived. Internet users are mentioned later in the paper as having a "vague sense of genres". Lee and Myaeng (2002) omitted mentioning users, and instead stated that genre is an "important property of text" (p.145).

I speculate that the lack of explicitly defined user groups is not seen as a glaring omission because of the broadness of the proposed genres: article, meeting minutes, course descriptions, home pages (Crowston & Williams, 2000), searchable indices, reports, FAQs, (Dewe, et al., 1998; Karlgren, et al., 1998), and research articles, editorials, reviews, etc. (Lee & Myaeng, 2002). The genres are broad enough that it could just be assumed that most everyone is familiar with these genres. We also observed this phenomenon with Ferguson's (1994) and Biber's (1994) definitions of genre reviewed earlier.

Although no genres are proposed, (Nilan, et al., 2001) is the only study that explicitly defines its user group: web users (though one might question the degree of explicitness in this definition). Haas & Grams (1998a) incurred no such definitional burdens of genre in that they did not claim that their categories are genres. They are "page types" classified on "functional, genre-related factors". Interestingly, one of their classification characteristics is mentioned as "intended audience". Of the seven page types proposed, they mentioned audience for only one type: the home page. The other six types are so broad (e.g., documentation, text, and multimedia) that no specific audience can be attributed.

As in the traditional (i.e., non-web) genre literature reviewed earlier (e.g., Ferguson, 1994), web genre studies such as these can afford to omit specific discussions of user groups by virtue of the broadness of their proposed genres. Lack of theoretical completeness in these cases does not cause big problems. However, it is unclear whether broad genres alone (or any genres) will be useful for web retrieval.

A well-defined user group is important in the definition of web genre because it distinguishes a web genre from a web page type. Admittedly, the distinction is fuzzy. Theoretically, two or three people can constitute a group. If they communicate distinctively via web pages in recurring situations amongst themselves, then strictly speaking, this is a web genre. One way to think about it is: the larger the number of people that understand the communicative class, the more like a genre (and the less like a type) it is. Another way to make the distinction is to say that a type has no user group defined. A web genre without a user group is not a web genre. It is a web page type. This is a stricter definition of web genre than the previous one. Most of the studies reviewed here (that purport to study web genres) propose genres that fail to meet this definition. In summary, web genre is a subordinate concept to web page type. A web genre is a type of web page that is recognized by a specific user group.

We can make some generalizations about web classifications, and whether they should be referred to as web genres or web page types. If the categories are developed *a priori*, without user input (Crowston & Williams, 2000; Haas & Grams, 1998a; Rehm, 2002; Lee & Myaeng, 2002; Björneborn, 2004), the classes are most likely web page types, unless some evidence is provided that a specified user group can recognize them (e.g., experimental evidence indicating user familiarity (Dillon & Gushrowski), or the classes are consistent with

"socially recognized" schemes (possibly like the Art and Architecture Thesaurus (Petersen, 1994) as in (Crowston & Williams, 2000)), or the "it's just plain common sense" argument. If the categories are developed *a posteriori* with user input (e.g., Nilan, et al., 2001; Dewe, et al., 1998; Stein & Meyer zu Eissen, 2004), the categories are more likely to be web genres, with user groups based on the experiment participants who provided the data.

In an overview of the literature on uses and characteristics of genres, Beghtol (2000) noted that, ideally, genres should be mutually exclusive, i.e., each web page is assigned to one and only one genre. She also noted that this ideal may not be possible "in any classification or in any domain". The situation of classifying texts by purpose, when some texts have multiple purposes, was given as an example. A text might also benefit from multiple genre classification, if it contains multiple genres. Haas & Grams (1998a) encountered this exact phenomenon during their classification efforts. And, as mentioned earlier, a genre may contain another, e.g., a newsletter that contains an events calendar, a situation possibly best described by multiple genre designators. This circumstance has been labeled as "embedded genres" (Crowston & Williams, 2000). As noted earlier, the sociolinguist, Biber, provided a facet for this in his taxonomic framework (Biber, 1994).

Mutual exclusivity of genres is not mentioned per se, but its use is implied in four studies (Crowston & Williams, 2000; Roussinov, et al., 2001; Rehm, 2002; Lee & Myaeng, 2002). Only one study reviewed here explicitly required mutual exclusivity of its genre palette (Nilan, et al., 2001). In fact, the inability to reconcile this requirement with the overlapping genre terms of the study participants led the authors to abandon the analysis of the data set. Furthermore, the authors questioned whether achieving mutually exclusive

genres developed from users' terminology was possible. And, even if it were, could users distinguish between the genres?

Haas & Grams (1998a, 2000) proposed a solution to the problem. Drawing on Biber's (1994) concept of embedded text from differing registers (genres), the page types would be considered to be "primitive building blocks that may occur singly or in combination on a page." In this scenario, pages could be classified into one or more "building block" classes. Rehm (2002) incorporated this concept into his "web page modules". Finally, Karlgren, et al. (1998) felt that allowing pages to be classified into more than one genre would improve the accuracy of the automatic classification process.

Web classifications may also vary in terms of the unit of analysis. Most of the studies reviewed here consider the individual web page to the primary instance of type or genre. The exception is Rehm's (2002) "web genre types", which may consist of multiple pages. Crowston & Williams (2000) also noted, for example, that web documents, such as an FAQ, can be split into several small pages connected by hyperlinks. Haas & Grams (1998b) proposed a link classification that could be used in a web page classification. In other work, Shepherd & Watters' (1998, 1999) study of "cybergenres" resulted in six, broadly categorized types of *web sites* (home page, brochure, resource, catalogue, search engine, and game), which also consisted of multiple pages (a single page, and all the pages it linked to). Finally, consistent with the idea of "embedded genres", multiple genres can exist on one web page (Rehm, 2002).

Another characteristic of genre palettes is their intended degrees of exhaustiveness. Is the palette intended to include all webpage instances into a genre, or are there genres (identified or potential) that are deliberately excluded from the classification? For example,

the authors of the studies reviewed here whose palettes were based on random samples (Crowston & Williams, 2000; Haas & Grams, 1998a) did not claim exhaustiveness. Instances of genres not included in the random samples were not considered.

In (Dewe, et al., 1998; Karlgren, et al., 1998), it is unclear whether the palette was intended to be exhaustive. The authors collected over 1800 sample pages, but only present categorization data for 1,358. The reason for the difference is unreported.

Two exploratory studies (Roussinov, et al., 2001; Nilan, et al., 2001) stressed that their genre palettes were not intended to be exhaustive. Their goals were to identify the genres that are useful to searchers. By asking searchers to identify the genres of pages retrieved while they are engaged in actual searches, the authors aimed to identify the genres that were most useful for various search purposes. Exhaustiveness in this case is unnecessary. There is no need to include genres that are not helpful in improving the search process. Roussinov, et al. (2001) presented a summary of most frequently reported search purposes, that includes the genres most relevant to each purpose.

Exhaustiveness of the genre palette may actually be detrimental to its viability for improving search. We have noted that the web is obviously a blend of user groups. Thus, web users may have difficulty recognizing genres used primarily in user groups other than their own. This would argue against developing an exhaustive genre palette based on the entire web. Nilan, et al., (2001) concluded in their study that “conceptualizing web users as a discourse community is simply too broad to be effective.” They went on to speculate that their future research efforts would concentrate on smaller, more focused discourse communities. Independently, Rehm (2002) criticized the “Internet-wide random sample generation” methods (like that of Crowston & Williams, 2000; and Haas & Grams, 1998a) as

"inherently leading to results which are too broad, and rather vague." To overcome this problem, Rehm restricted the domain of his genre palette to German universities. He reasoned that most universities are similarly structured, and thematically focused on research, education, and administration. Thus, the restricted domain could lend itself to a more effective genre palette, but only for a more restricted user group.

One last point regarding exhaustiveness is the difference between intended exhaustiveness and actual exhaustiveness. Some pages may not fit well into any of a palette's genres. Studies reviewed here, which reported this data, reported significant percentages (12-32%) of unclassifiable pages, that is, pages on which coders did not agree.

Other Uses of Web Genres

Web genres can be useful in endeavors other than retrieval by genre. For example, genres can augment subject-oriented retrieval. Chen, et al. (1996) found that "personal home pages" (which discuss the professional experience, education and personal interests of their owners) were difficult to classify by subject. Instead, they suggested a separate personal home pages category. They speculated that searching by subject would be more effective if the personal home pages could be removed prior to the subject categorization process (although results reported later in this dissertation suggest that doing so may cause useful information to be missed).

Development of genre palettes could aid manual categorization efforts. For example, the Dublin Core Metadata Initiative (http://dublincore.org) recommends, as part of its metadata element set, that a resource-type descriptor be used to identify the genre of a resource. Thus, a genre palette could provide the possible set of values for this descriptor in a

manual categorization of web pages. Also regarding the Dublin Core, Greenberg (2004) suggests that a genre descriptor could be used to determine the appropriate extraction algorithm for automatic metadata generation.

It has been suggested that genre identification could also benefit browsing the web. In their study on browsing with the aid of keyword-based user profiles, Shepherd, et al. (2001a, 2001b) suggested that profiles, containing characteristics such as genre, may be more effective than those profiles based on content alone.

An entire palette of web genres constitutes a typology, which can be useful in conducting research. For example, the webpage typology of Haas & Grams (1998a; 2000) was used to help select what kinds of pages to present to subjects in Kelly (2004), and also to discover differing rates of web page content change in various types of pages (Tan, et al., 2001). It has also been suggested that a webpage typology would be useful for solving in issues in cultural asset management (Ross, 2001).

Finally, genres can provide guidelines for web authors and designers. Authors and designers could use genres to easily become familiar with reader expectations of a particular genre. For example, genres may have varying expectations of functionality (Shepherd & Watters, 1999). They could even choose the genre of a proposed webpage based on the genre that most appropriately matches the developers' communicative intent (Crowston & Williams, 2000; Haas & Grams, 1998a).

In summary, the successful application of genre theory to web retrieval is just beginning to be investigated. Many open issues in genre palette development remain such as the genres' level of abstraction, the mutual exclusivity of genres, the exhaustiveness of the genre palette, and the usefulness of particular genres for searching.

In the next section, we look at how genre theory relates to the relevance literature in information retrieval, and what the implications are for retrieval by genre.

Genre and Relevance in Information Retrieval

Researchers have investigated the application of genre theory to information retrieval (e.g., Nilan, et al., 2001; Karlgren, et al., 1998). The goal was to give the user "a better chance of finding what he/she needs". (Nilan et al., 2001, p. 330). How can we expect genre to help? How is genre related to relevance? Genres are essentially categories. People make sense of the world by categorizing things (Smith & Medin, 1981). This process of categorizing allows people to make inferences about these categorized objects. These inferences help people determine the objects' relevance to them. Given that genres are categories, it is hypothesized that the inferences made about a document based on its genre can help a searcher determine the relevance of the document to his/her information need.

This section begins by discussing how people's knowledge of categories makes topic-oriented retrieval possible, and proposes that a similar relationship exists for genre-oriented retrieval. The basis for this relationship, shared genre knowledge, is discussed next. Shared genre knowledge is the assumption that people have similar conceptions of what specific genres are. Three historical views of relevance in information retrieval (IR) are then reviewed, and their relationship to shared genre knowledge is discussed.

Shared Concepts

Understanding a concept gives one the ability to categorize novel entities as illustrations of that concept (Smith & Medin, 1981). To the extent that people share concepts, the concepts can be useful for IR. For example, because indexers and users share conceptual knowledge of

index terms, the index terms are useful for retrieval of specific subject-related material. Users understand what type of subject material is indicated by a specific index term (more or less[3]). Thus, IR with systems using controlled vocabularies is made possible by the shared knowledge of users and indexers.

Shared Genre Knowledge

Like knowledge of concepts, genres can be thought of as shared social constructs. For example, we take for granted that most everyone knows what to expect from business memos, personal letters, TV sitcoms, operas, etc. (in terms of general form, content, purpose, etc.). Recognition of genre depends on shared knowledge of genre characteristics between the readers/viewers and writers/producers. This shared knowledge can also be referred to as background knowledge, common ground, mutual knowledge and assumed familiarity (Giltrow, 1994).

To the extent that shared knowledge of genres exists, information seekers will understand what types of documents can be retrieved by a genre-augmented IR system. For a given genre, users can infer certain characteristics about a document of that genre. A hypothesis of this research is that these inferences can improve the accuracy of users' relevance assessments of documents. Although specification of topic is necessary, the indication of genre may be a useful addition to the facilities available to users for query specification.

This idea that there is some shared (or "public") knowledge with which people can make similar judgments can be related to what Schamber (1994) calls the "information view" of relevance. "The information view assumes that nonusers can adequately predict whether

[3] Problems can arise when users specify different terms than indexers to indicate similar subject matter (see, for example, Furnas, et al., (1987) and Solomon (1992).

certain information will solve users' information problems" (p. 7). People other than those with the information need, can determine relevance. The information view contrasts with two other views: the system view and the situation view. In the system view, the system determines the relevance of a document by matching the query representation with the document representation. Relevance judgments between specific document-query pairs are assumed, and system performance is measured by how closely the system's results match these pairings. This view is taken by studies usually concerned with improving the system aspects of information retrieval such as document and query representations (e.g., Cleverdon, (1997) and Salton & Buckley (1988)). In the situation view, the relevance judgment can only be made by the end-user of the system. Relevance is an individual, personal decision in this view. All of the user-defined relevance criteria studies reviewed in this paper come from this perspective (Schamber (1991), Park (1992), Cool, et al., (1993), Barry (1994), Wang & White (1995), Schamber & Bateman (1996), Bateman (1998), Spink, et al., (1998), Tang & Solomon (1998), and Maglaughlin & Sonnenwald (2002)). An example of the contrast between the system view and situation view is the difference between the IR techniques of automatic query expansion (Buckley, et al., 1995) and relevance feedback (Salton & Buckley, 1990). In automatic query expansion, the additional terms used to automatically expand the query come from the top-rated relevant documents, as *determined by the system* (and the designers of its specific algorithms). In relevance feedback, the documents used as the source of additional terms are those that are *judged as relevant by the individual user*.

For our purposes, the important distinction between the three views (system, information and situation) is who (or what, in the case of the system view) is making the relevance judgment. In the system view, relevance is determined by the system. In the

situation view, only the person with the information need can determine relevance. In the information view, anyone (with sufficient public knowledge) can judge relevance. References to research grouped by these three views of relevance can be found in Schamber (1994). At this point, discussion of the system view is suspended for now, as the system view is irrelevant to the human process of making relevance judgments.

The distinction between shared (public/objective) knowledge and individual (private/subjective) knowledge is also discussed in (Foskett, 1972; Kemp, 1974; Saracevic, 1975; and Lancaster & Warner (1993)). Foskett's (1972) and Kemp's (1974) notion of "relevance" is associated with relevance judgments based on public knowledge (i.e., objective). (This association of the "objective" with "public knowledge" does not refer to its validity. Rather, public knowledge is objective in the sense that people have the same information (i.e., shared knowledge).

In contrast with Foskett and Kemp's "relevance", their notion of "pertinence" can be seen as relevance based on private knowledge (i.e., subjective). Saracevic (1975) notes that numerous authors have made this or similar distinctions (e.g., Rees & Saracevic, 1963). Lancaster and Warner (1993) also use the terms relevance and pertinence similarly, but state that they do not completely agree with Foskett and Kemp about the objectivity of relevance judgments. They observe that disagreement between judges making relevance judgments is not uncommon. (See also Saracevic, 1991). For them, relevance indicates the relationship between a document and a query in the eyes of a judge (Schamber's (1994) information view), and pertinence is the relationship between a document and an individual's information need (Schamber's (1994) situation view). However, Lancaster and Warner (1993) point out

that inconsistency between individual judges may have no significant effect on some IR system evaluation measures (Lesk & Salton, 1968).

More recent support for the concepts of shared and individual knowledge comes from Hjørland's "socio-cognitive perspective" on information science (Hjørland, 1995; 2002) in which he asserts that "subjective and objective relevance criteria develop during individual and collective cognitive development" (Hjørland, 2002, p. 210). However, not all recent IR models of relevance employ this distinction. In regard to relevance judgments made by judges, Borlund (2003b) writes:

> It is this kind of subjective relevance assessment we subscribe to individual assessors who participate in common IR experiments like TREC, although the judgments are traditionally intended to be of an objective nature. (p. 915)

Borlund calls this type of subjective relevance, "intellectual topicality" (thus distinguishing it from the "topicality" of the system view (Schamber, 1994)). There is no mention of the shared/individual knowledge distinction. Interestingly, Cosjin and Ingwersen (2000) hinted at a resolution of the discrepancy in this objective vs. subjective debate with their proposal of another type of relevance: socio-cognitive relevance.

> We regard socio-cognitive relevance as a subjective type of relevance determined by the individual actor in interaction with other actors within a community. When tangible and measured, it may often exhibit statistically objective characteristics (inter-subjectivity). (p. 546)

Thus, according to Cosjin and Ingwersen (2000), relevance assessments based on shared knowledge are subjective in nature, but may have a high degree of agreement. Table 2.2 summarizes how the literature addresses the concepts of shared versus individual knowledge.

Table 2.2 Concepts related to the distinction between shared and individual knowledge

Literature	Shared Knowledge	Individual Knowledge
Foskett (1972); Kemp (1974)	Relevance is objective and public.	Pertinence is subjective and private.
Lancaster & Warner (1993)	Relevance can be determined by a judge.	Pertinence is determined by an individual's information need.
Schamber (1994)	"information view"	"situation view"
Saracevic (1975)	"subject knowledge view"	"pertinence or destination knowledge's view"

I suggest that genres can aid retrieval by tapping in to users' store of shared genre knowledge. Users' genre knowledge could allow a fuller expression of their needs to the system (Taylor's (1968) "compromised need"). Also, query results could be labeled by genre. This would allow the users' genre knowledge to supplement the purely semantic content of the document (or document surrogate), giving the users additional clues to the document's relevance or non-relevance.

Implications of Relevance Criteria for Retrieval by Genre

Since the early 1990's, a flurry of experimental studies that identify "user-defined relevance criteria" have appeared in the information science literature. Despite the diverse combinations of tasks, settings, methodologies, user populations, etc., significant overlaps in identified criteria have been observed (e.g., Barry & Schamber, 1998; Maglaughlin &

Sonnenwald, 2002). Thus, it seems reasonable to believe that these criteria can be considered legitimate bases on which users make relevance judgments (in some situations, at least).

Regarding the current proposal for genre-based retrieval, the success of the technique rests on the users' abilities to relate genre to their information needs. It seems plausible that, if users can associate specific relevance criteria (for their particular information needs) with specific genres, then document selection by genre could facilitate the retrieval of relevant documents. For example, if a searcher is looking for the date of an upcoming social event or community meeting, then restricting the search results to the genre of "calendar" can be a potentially powerful option, by helping avoid unwanted results like information about the organization's mission, its officers, last year's picnic, etc.

In the next section, we explore the process by which people classify things, and how that might affect people's genre recognition, and the usefulness of specific genres for the search problem.

Cognitive Factors in Retrieval by Genre

Genre is a socially recognized document type. In order for genre-augmented queries to improve web search, the user must be able to recognize the genres that are available for selection or deselection in the query specification process. Knowledge of how users might classify documents into genres can help guide the construction of the genre palette for the user interface. This section reviews literature on human categorization behavior, and applies it to the problem of genre-augmented web retrieval.

Prototypes and Categorization by Family Resemblance

People tend to view categories in terms of clear cases of category membership. For example, in the U.S., a robin seems to be an ideal example of the category of birds (e.g., more so than a penguin). These clear cases are called prototypes. Prototypes "contain the attributes most representative of items inside and least representative of items outside the category" (Rosch, 1978, p. 30). They are "category members that have a special cognitive status – that of being a 'best example' " (Lakoff, 1987, p. 41).

Research has shown that people who share a worldview or life experiences, generally agree on how typical a category member is for a given category. As previously stated, most people would agree that robins are better overall representatives of the bird category than penguins, and even chickens or ostriches. This phenomenon is empirically well-documented (Rosch, 1978). It has also been shown that "the more prototypical of a category a member is rated, the more attributes it has in common with other members of the category and the fewer attributes in common with members of the contrasting categories"[4] (Rosch, 1978, p. 37).

These findings support Wittgenstein's (1953) hypothesis that people categorize according to "family resemblances" rather than by defining necessary and sufficient conditions (Lakoff, 1987). For example, ostriches are seen as birds because of the (admittedly vague) "resemblance" they have to prototypical birds. However, a definition of bird such as "having wings and can fly" (a definition with which most people would agree), fails to account for ostriches as birds.

[4] Obvious exceptions to this phenomenon are goal-oriented, artificial classes named by Barsalou (1983) as "ad hoc categories".

A result of categorization by family resemblance is the potential for categories with fuzzy or non-rigid boundaries (e.g., the category of "tall men") (Wittgenstein, 1953, Rosch, 1978, Lakoff, 1987).

The Basic Level of Categorization

Like genres, any concrete or abstract object can be indicated at varying levels of abstraction. For example, a particular pet may be referred to as a "dog", "poodle", "mammal" or "animal". It is possible for a specific object to be all of those things. Research has shown that humans have a preferred level of categorization that has become known as the "basic level" (Rosch, 1978). In the preceding "pet" example, "dog" is the basic level category term, whereas "poodle" is at the lower-level, more detailed subordinate level, and "mammal" is the more general, more abstract superordinate term for "dog". ("Animal" is an even more general, more abstract superordinate term for "mammal").

Lakoff (1987) summarizes the work on basic level categorization done by Rosch, Berlin (Berlin, et al., 1974) and others. According to Lakoff, the following is true of the basic level:

- People name things more readily at this level.
- Languages have simpler names for things at this level.
- Categories at this level have greater cultural significance.
- Things are remembered more readily at this level. (Lakoff, p.33)

Rosch and her colleagues observed that "basic level distinctions…are characterized by overall shape and motor interaction and are at the general level at which one can form a mental image" (Lakoff, p.49). For example, in general, humans can easily associate images and physical actions with basic level categories. One can picture a chair and know how to sit on it or get up from it. The superordinate category to "chair", "furniture", does not have such

natural associations. Rather than associate images and actions with the overall category "furniture", one thinks instead of types of furniture (i.e., table, chair – basic level categories). Lakoff summarizes the concept of basic level categories as follows:

> Perhaps the best way of thinking about basic-level categories is that they are "human-sized". They depend not on objects themselves, independent of people, but on the way that people interact with objects: the way they perceive them, image them, organize information about them, and behave toward them with their bodies. The relevant properties clustering together to define such categories are not inherent to the objects, but are interactional properties, having to do with the way people interact with objects. (Lakoff, p. 51)

People find basic level categories easier to use than superordinate or subordinate categories.

> Superordinates have only a few features in common, and these tend to be abstract, functional properties. Subordinates have many features in common, but most are also properties of their basic category; their novel properties tend to be minor perceptual or functional modifications of typical properties at the basic level. *These rather different category structures have a similar effect of making the two levels more difficult to use…Speakers tend to avoid describing individual objects with names at either level* [italics added] (Murphy & Lassaline, 1997, pp. 114-115).

Although basic level categories tend to be shared by people, differences in groups' basic level of categorization have been observed, and can be explained by various factors. Expertise in a domain can affect the preferred level of categorization. Experts have been shown to use a higher percentage of subordinate categories in their domain of expertise than in a domain in which they lack expertise (Murphy & Lassaline, 1997). One possible interpretation is that experts' greater knowledge of a domain makes them more aware of distinctions between domain objects than novices. Other research on the effect of expertise on categorization has shown that experts sometimes group things together that novices tend to classify separately (Murphey & Lassaline, 1997). Presumably, domain knowledge sometimes allows experts to see similarities that novices do not.

Preferred levels of categorization may also be affected by culture. For example, Mayan children have been observed to consistently use plant names at a lower level than children in Berkeley, California (Murphey & Lassaline, 1997). The implication is that urban dwellers do not develop as elaborate biological distinctions as those in non-industrialized societies. Lakoff (1987) also suggests expertise and culture as factors affecting levels of categorization.

Certainly, the context of the categorization act affects the level used. For example, someone who is about to buy a chair while standing in the middle of a furniture store "will think and speak about chairs at other than the basic level of 'chair'"(Rosch, 1978, p. 42). Rosch justifies the derivation of experimental findings on categorization from artificial laboratory settings by proposing the existence of a default context for objects. In other words, if people are not given a context for an object, they will assume the context of the situation in which an object is typically used.

In summary, people tend to categorize things by family resemblance, people have a preferred level of categorization, and the preferred level is influenced by expertise, culture and context.

Genre and Cognitive Factors

Paltridge (1995) proposed a model of genre based on the prototypicality theory documented by Rosch (1978). Genre instances are classified into genres based on family resemblance. Atypical genre instances (those with little or no typical characteristics of form) are classified on the basis of pragmatics. This is consistent with the definition of genre proposed in an earlier section of this literature review. A key part of that definition of genre

is the user group. In Paltridge's model, this notion is subsumed under the aspect "Institutional Understandings", which includes:

> "…the ideology and beliefs of the particular discourse community as well as the shared understanding of how the particular text should proceed [and] an understanding of what is to be assumed to be generally accepted as common knowledge within the particular area…" (Paltridge, 1995, p. 401).

Paltridge's model does not specifically address the differing levels of abstraction of genres.

Steen (1999) also relates prototypicality theory to genre classification, postulating that the level of genre is the basic level, and that super- and sub-genres are at the superordinate and subordinate levels of classification, respectively. Since discourse participants are responsible for the production and interpretation of texts, Steen favors what was termed in an earlier section as a user-driven palette. "A generally valid taxonomy of discourse should not project our expert scientific view of discourse types onto the range of discourse but instead begin with an examination of discourse concepts as they are valid for ordinary language users." (p.111) Nilan, et al. (2001) made a similar case for "bottom-up" classification in their web genre study.

In the process of labeling 4,124 British National Corpus (BNC) files by genre, Lee (2001) applied prototypicality theory to the problem of determining the appropriate levels of abstraction for genres. He concluded that the level of abstraction does not matter, as long as it is useful. Research studies concerned with the comprehension of digital documents have confirmed many of the findings on human categorization behavior discussed here. Toms, et al. (1999) compared subjects' abilities to recognize prototypes of specific genres of digital documents, in three versions. The first version was the original document (used as the control in the experiment). The second version contained the text of the original document with all of the formatting removed, i.e., one big paragraph of text, i.e., content without form. The third

version retained the form of the original but all alphabetic characters were changed to ‘X’, and all numeric characters were changed to ‘9’, i.e., the original form, but without content. They found that, in approximately 1/3 of the cases, the genres of the form-only versions were identified correctly. Overall, the genre of the form-only version was also recognized 1/3 to 2/3 faster than that of the original, which is consistent with Rosch’s (1978) findings. Family resemblance (i.e., characteristic features of form) seems to be a strong factor in the categorization of digital documents. Curiously, the researchers did not find an expected effect based on expertise. They speculated that the experiment’s “novices” had more genre knowledge than originally estimated.

Dillon and colleagues (Dillon, 1991; Dillon & Schaap, 1996; Dillon & Vaughn, 1997; Vaughn & Dillon, 1998; Dillon, 2000) researched subjects’ abilities to recognize the text structure of articles of academic journals, and coined the phrase “the shape of information”, (a genre-esque combination of syntax and meaning) to describe how digital information is presented. In one study (Dillon, 1991), experts were able to recognize (in both paper and electronic formats) the sections of an “academic article” (introduction, method, results, discussion) to which individual paragraphs belonged, at average levels of 80% or better. Another experiment showed that novices made more errors in section recognition than experts (Dillon & Schaap, 1996). The implications of the studies are that shared knowledge of the academic article genre allowed subjects to recognize parts of the genre. Further, the expertise effect implies that experts have greater facility with a genre than novices do. This supports the idea that the user group is a key aspect of a genre’s definition.

Dillon & Gushrowski (2000) attempted to identify a specific web genre: the personal homepage. They sampled personal homepage sites, identifying and counting web page

elements such as title, author's email address, back to top button, etc. From the list of element frequencies, they created a set of sample pages that represented a continuum of prototypicality: pages with varying degrees of common and uncommon page elements. Their subjects were asked to rank the sample pages in order of how representative the pages were of typical personal homepages. The subjects' rankings agreed with the researchers' tally of typical personal homepage elements. In a second task, subjects' choices of typical personal page elements (selected from a list) were highly correlated with common elements from the original sample of personal web pages. The results of both tasks combined suggest that there is broad agreement of what a personal web page is and what features it contains. In terms of defining the user group for the genre, the authors noted that the subjects (57 graduate students in Information Science) may not be representative of web users as a whole. The implication, however, is that this is a "web wide" genre: "the shared matching of expectation and preference across this community of users suggests the personal home page might be the first unique digital information genre." (p. 205) Due to the authors' attention to the social recognition aspect of genre (e.g., "Genre conventions emerge across discourse communities over time to support the communication of ideas and information in socially and cognitively compatible forms", p. 202), their use of the term "genre" is consistent with the concept developed in this paper.

Few studies have been reported which measure user agreement regarding the genres of webpages. Kraft and Stata (2003) evaluated the efficiency of a system for identifying "buying guides" on the web. Seven evaluators from "diverse" backgrounds individually rated 20 web pages in each of 10 product categories, as to whether or not the pages were of the

buying guides genre. The evaluators received detailed instructions regarding the nature of a buying guide. For example,

> A buying guide is defined in terms of its intent. A buying guide is meant to help people at a certain point in the buying process. Imagine that you know nothing about digital cameras but you think you might want one. At the very beginning, you're less interested in the specific details of particular products and more interested in learning about the entire category. ... When looking at a document, ask yourself: Is this document useful given that I know little about this category and I'm trying to learn about it? (p. 90)

With the instructions, overall agreement among the evaluators was reported at 85%. Of course, the need for such extensive instructions could make one somewhat skeptical about whether "buying guides" is truly a web genre as defined earlier.

As one small part of a larger exploratory study, Santini (2005) instructed six participants to separate six web pages into 2 to 5 non-topical groups, and to assign labels using their own terminology to each group. The experimenter's labels and users' labels are reproduced in Table 2.3.

Table 2.3 Non-topical Labels assigned to web pages from (Santini, 2005)

Author	Study Participants					
	1	2	3	4	5	6
sermon	debate or controversy	short paragraphs	philosofical [sic] dissertation	analysis	editorial	commentary
e-shop	search/user's interface	links	e-commerce portals	merchant page	information index	selling
editorial	debate or controversy	long paragraphs	philosofical dissertation	analysis	editorial	commentary
introduction	news	short paragraphs	organization's Information	announcement/ press release	organization introduction	introduction
news release	news	long paragraphs	business Information	announcement/ press release	news	fact
search page	search/user's interface	user input	e-commerce portals	search page	information index	description search

Despite the relatively loose criteria given to the subjects, some agreement on conceptual genres can be discerned. Santini reported that future work would include a more extensive study of user agreement.

Implications for Retrieval by Genre

What we know about human categorization behavior suggests that genre-augmented retrieval can potentially be successful. The fact that people naturally group things together on the basis of family resemblances indicates that genres, which typically have characteristic forms, may be recognized easily. Given that genre recognition is a form of categorization, and, that people tend to name things at the same preferred levels of categorization, there can be widespread agreement on what is and is not an instance of a specific genre. Thus, it may be possible for searchers to select documents based on recognized genres. Also, the fact that shared preferred levels of categorization are defined by interactional properties, is consistent with genre definitions based on purpose (Swales, 1990) or action (Miller, 1984). People name things according to how they typically interact with them. Genres are names for documents commonly associated with people's specific purposes and actions.

The literature reviewed here on human categorization behavior has many implications for the construction of a genre palette. Because categorization by family resemblance can result in categories with fuzzy boundaries, it may be worthwhile to allow genre instances to belong to more than one category (i.e., no requirement of mutually exclusive categories) -- if the additional system complexity is manageable. Also, genres should be at the preferred level of categorization, since that is the most frequently used level. It can be affected by people's culture and expertise. Thus, genres whose user groups have significant differences in cultural or expert qualities from the other genres in the palette, may need to be avoided, if the

retrieval system is intended to support a broad set of user groups. In that case, genres with high levels of abstraction are needed. Beghtol's (2000) call for "culture-neutral methods for classifying entities" indicates a similar concern over the relationship of genre and the globalization of information. Yates & Orlikowski (1992) posited a similar relationship between a genre's level of abstraction and the breadth of the community that recognizes the genre. Also, studies of the classification of web pages have found that agreement about a document's genre is more difficult at lower levels of abstraction (e.g., Haas & Grams, 1998a).

Finally, categorization is context-related. Genres listed in a search engine's user interface are inherently out-of-context. So, presumably, users will provide their own notions of the genres' typical usage as a default context.

There are many different ways in which genre could be used to improve information retrieval. For example, searchers could be allowed to include or exclude certain genres of documents from their search results. Genre information could be used implicitly by systems to assist in automatic query reformulation through relevance feedback mechanisms. Also, search results could be explicitly annotated with document genre. This could help searchers make better judgments regarding which pages to browse, or it could help by giving searchers genre information which they could use in their own query reformulation.

This dissertation explores this last idea in depth. (See Chapter Three for the specific research questions, and Chapter Six for the description of the specific study.) The next section reviews the literature regarding the evaluation of document surrogates.

Evaluation of Document Surrogates

Little work is known that compares web searchers' relevance judgments of web pages with their associated summaries in order to measure the summaries' effectiveness. Lan (2002) compared graduate student researchers' qualitative assessments of webpage/summary pairs. He found that 59% of the paired judgments were dissimilar. However, only one study other than this current work is known to have compared quantitative assessments of web pages and their surrogates (Tuffs, 2002; Shou, et al., 2003). A change in judgment from surrogate to full-text was reported to only have occurred 43.5% of the time, on a less complex task than that in Lan's research: finding a review of a musical band's new album. However, unlike the current work, this study compared accuracy and timing of surrogate judgments across two systems (Google and SpeechBot – an audio file retrieval system), rather than across two different types of surrogates on the same system. Other than that, the study is very similar to a study in this current work, and will be discussed in detail in the Chapter 6, after the details of this work have been presented.

The paucity of this kind of research in the literature may be due to the long-term trend towards more qualitative research in systems evaluation, as discussed by Belkin and Muresan (2004). Recent work on search evaluation has focused on measuring a search's overall effects on the user, e.g. the information problem shift (e.g., Spink, 2002), and the value of the search results as a whole (e.g., Su, 2003).

A few studies have compared different types of web page summaries with each other, rather than measuring them against the assessment of the actual page. Amitay (& Paris, 2000; Amitay, 2001) had subjects select a search result from a list, view the corresponding webpage, and then, on a scale from one to seven, rate how happy they were with the result,

and how easy it was to find the needed information. Ratings of three different styles of search results were compared: query-biased (e.g., Google), a page's first few lines (e.g., Altavista), and a summary type dubbed "InCommonSense" that gleans summary sentences from pages that link to the page being described. Similarly, White, et al. (2003a,b) measured users' ratings of summaries following a web search task, without soliciting a comparable rating of the full-text. Instead, users rated perceived qualities of the summaries (i.e., relevance, importance, usefulness, and completeness), and those ratings were compared by the four summary types being studied.

With the exception of (McLellan, et al., 2001), most non-web comparisons of surrogate and full-text judgments are fairly dated (see Appendix A for a comparison with other such studies). Some relatively recent non-web summary comparison studies have compared a user's topical assessment of a summary with a group of experts' assessment of the full-text (e.g., Mani, et al., 1999, Tombros, 1997; see Appendix A). In the other three studies summarized in Appendix A (Kent, et al., 1967; Saracevic, 1969; Marcus, et al., 1978), participants judge multiple surrogate types and the full-text of each document. In the current research, each participant judges only one (of two possible) surrogate types, and the webpage. Thus, this dissertation's study differs from the others in that user performance measures based on multiple surrogate types (genre-annotated, and not annotated) are compared across participants rather within participants.

One well-known older study of document surrogates, Janes (1991a), compared the changes in relevance judgments as different surrogate types were introduced to the user. However, no judgments of the full-text documents were solicited.

One reason for the lack of comparable web studies may be the low cost of evaluating a non-relevant web page, as compared with the effort needed to acquire traditional printed materials. Another reason may be the relatively more time-consuming nature of evaluating web pages, as compared with the relatively quicker evaluation time of search results. Researchers wishing to evaluate summaries can gather more assessments of summaries, if full-page assessments aren't required by the experimental design. Finally, it could be thought that the poor quality of web search engine summaries is so well-known, that without any feasible proposals for improvement, any critical evaluation would be superfluous.

CHAPTER THREE

RESEARCH QUESTIONS

In theory, the use of document genre as a descriptor in the document representation of a web IR system seems promising, yet the original question posed in the Introduction (Chapter One) remains unanswered. What document descriptors can be incorporated into an IR system's document representation in order to improve search effectiveness? The precise values of the descriptor have not been specified: what genres would be useful for improving the search effectiveness of a web search engine? The following three criteria for a genre's use as a document descriptor are hypothesized:

1. The users of the system must possess sufficient knowledge of the genre. A child interested in finding out how to make cupcakes, who does not know what a "recipe" is, cannot use that genre as a relevance indicator. A genre must be recognizable to the searcher. Using Yates and Orlikowski's (1992) concept of normative scope, the genre's normative scope must be large enough to include (at least some of) the users of the system.
2. A searcher must be able to relate the genre to his information need. Genres that are orthogonal to the user's need may not be useful. For example, consider the task of someone who wants to find out the locations of local companies. It is unclear how helpful the use of the "resume" genre would be. Certainly, resumes could be used to

identify local companies, but more complete information might be much more easily found on other types of pages. Thus, neither searching only resumes, nor excluding resumes from the search would seem like a profitable search tactic. A genre must be useful to the searcher, in relation to the information need.

3. As noted in the discussion of non-topical descriptors, these descriptors are not explicitly contained in documents, and must be inferred and assigned to documents by an automated process, for practical use. Thus, a genre must be predictable by a machine-applied algorithm.

In summary, a good genre candidate for document descriptor should be recognizable to searchers, useful for searchers' information needs, and predictable by machine algorithm. The exploration of machine algorithms for prediction of web page genre has been left for future work.

The Research Questions

The main research question for this dissertation was: can a genre palette be developed that is both recognizable and useful for web searching? To answer this question, work was divided into four steps, each with its own sub-questions:

1. What genres of web pages do users perceive? How would they name and define them?
   (Chapter Four)

2. What level of agreement between participants is possible in classifying web pages using genre names and definitions derived from step #1? Which genres among those with similar names and definitions will participants tend to use? Is it possible to produce a palette of genres that users can assign to webpages with an acceptable level of agreement?
   (Chapter Four)

3. Given a genre palette created as a result of step #2, what level of agreement between participants is possible in classifying a different set of web pages? (Chapter Five)

4. Assuming that a genre palette produced in step #2 is validated by step #3, will people be able to make better relevance judgments of search results using document descriptions that include genre? Will people be able to make faster relevance judgments of search results using document description that include genre? Will users perceive document descriptions that include genre as an improvement over those that do not? (Chapter Six)

To increase the chance of developing a recognizable palette, it was decided to restrict the domain of the pages covered by the palette (hence, restricting the user groups of the genres of the palette as well). Pages from the "edu" domain only were considered, as in (Rehm, 2002). Rehm criticized "Internet-wide random sample generation" methods as "inherently leading to results which are too broad, and rather vague." To overcome this problem, Rehm restricted the domain of his genre palette to German universities. He reasoned that most universities are similarly structured, and thematically focused on research, education, and administration. Thus, the restricted domain could lend itself to a more effective genre palette. Other studies also have suggested that a genre palette that does not cover the entire Web may be more conducive to producing an effective palette for use in the search process (e.g., Nilan, et al., 2001; Roussinov, et al., 2001).

It was decided to base the palette on user terminology, as a pilot study based on *a priori* categories produced a low level of user recognizability (Rosso, 2002), and as explained in Chapter 2, genres can be considered to be folk typologies, assigned on the basis of use (e.g., Biber, 1988).

Step #4, designed to address one aspect of genres' usefulness for searching, sought to determine whether describing documents by genre in the list of search results could improve

web search effectiveness. Thus, step #4 is concerned with the second point at which relevance criteria can impact the search process (see Figure 1.1).

Participants' relevance judgments of genre-annotated and un-annotated search results were compared based on the time required for participants to make relevance judgments, and the consistency of those judgments (when compared with the judgments of the corresponding web pages). Participants were also asked to report any preferences for the annotated or un-annotated style of search results. The null hypothesis for each question is that description by genre does not make a difference. Faster relevance judgments were measured by the time it took to judge a set of document descriptions. Relevance judgment consistency was measured by comparing the judgment of each search result with the judgment of the corresponding full-text document. User-perceived improvement was measured with qualitative survey questions.

Table 3.1 summarizes the methodology and results of the four parts of this research.

Limitations

The studies in this dissertation were designed as preliminary investigations into users' ability to recognize genres, and to use genres in making relevance judgments. Results may generalize only to similar populations and information tasks in the edu domain.

Ideally, users with their own real information needs would be used in the study of relevance judgments. However, in order to isolate the effect of genre on the relevance judgment process, it was decided not to have users formulate queries using genres, and conduct their own searches. It has been shown that simulated situations can be acceptable substitutes for users' real needs in experimental settings, e.g., (Borlund, 2000). The situations

used in this study were tailored to the users, as recommended by Borlund, and were pilot-tested.

Table 3.1 Overview of the studies undertaken in this dissertation

	Methodology	Product
May 2004 Survey of User Terminology (Chapter Four)	3 participants individually separated 100 webpage printouts into stacks according to genre, assigning names and definitions to each genre.	A collection of 48 genres names with definitions
July-August 2004 User-based Refinement of Terminology into a Tentative Genre Palette (Chapter Four)	10 participants individually classified 100 webpages (same as in the previous study) using the 48 genres (plus a "suggest your own") category.	A palette of 18 genre names and definitions
September 2004 User Validation of the Genre Palette (Chapter Five)	In an online experiment, 257 participants each classified a new set of 55 webpages using the 18-genre palette.	Validation of participants' ability to classify pages using the palette
November 2004 Measurement of User Relevance Judgments of Genre Annotated Search Results (Chapter Six)	32 participants performed 4 tasks. In each task, participants judged the usefulness of 20 search results & 20 webpages according to an assigned task scenario.	Comparison of participants' performance with and without genre annotated search results

This page left intentionally blank.

CHAPTER 4

DEVELOPMENT OF THE GENRE PALETTE

As noted in previous chapters, the web genres to be used in a search interface must be recognizable to the users, a precondition of usefulness for searching. This chapter details the development of the genre palette used in the study of the effects of genre-annotated search results on the relevance judgment process (Chapter Six). This palette development was conducted, and is described, in two phases (please refer to Table 3.1).

Survey of User Terminology

In the first phase, a card-sorting study was undertaken to inform the genre names and definitions used to describe those genres to be used in the palette. The goal was to obtain information on what genres users perceive in web pages. "Sorting is a conceptual mapping technique and so is appropriate when the goal is to discover categories that people use" (Rugg & McGeorge, 2003, p. 2702). (For further descriptions of the card-sorting methodology, see also (Rugg & McGeorge, 1997; and Jonassen, et al., 1993.)

Method

Participants were asked to separate a stack of printed web pages into piles of pages that they thought were of similar genre, according to the following definition:

> A document genre is a category of documents characterized by similarity of function, style, form or content. Traditional document genres include, for example, business letters, cooking recipes and greeting cards. Note that a document's genre is not the same as its subject. For example, a business letter may be about the availability of a new product, or an invitation to interview for a job. Both examples are business letters, but their subjects are different.

Printed (as opposed to on-screen) pages were used in order to make it easier to stack the pages.

A pseudo-random sample of 200 web pages from the edu domain was collected from search engine results from Google, produced by single-term queries, with each query consisting of one of the most frequently-used words in the English language and limited to the edu domain. The idea is that a high frequency word (like "she", for example) is not significantly related to the content or "kind" of page retrieved from the search engine. The list of high frequency words used was obtained from the Brown corpus (Francis & Kucera, 1982). (See Figure 4.1 for the algorithm for collecting the pages.)

Figure 4.1 Algorithm for collecting a pseudo-random sample of web pages

1. *Starting with the most frequently used word (Francis & Kucera, 1982), use the word as a single-term query to the search engine.*
2. *Download the pages (and associated files: images, etc.) that are the 200^{th,} 400th, 600th, and 800th results of the query. (If a page is unavailable, attempt to download the preceding result. For example, if #200 is unavailable, try #199 and so on.)*
3. *Repeat Steps 1 & 2 with the next most frequently used word until the sample is complete.*

Of the 200 pages collected using the algorithm, only 133 were unique. The occurrence of duplicates in the sample was caused by a misunderstanding of the conditions under which Google treats common words as "stop words", which essentially remove the term from the query. (Subsequent samples were collected using the same algorithm, except with the addition of a "+" in front of the search term, in order to assure the term is included in the query.)

The pages were printed in color using an inkjet printer on 8.5" by 11" paper. Page backgrounds were not printed because that inhibited the readability of many pages, as well as using a lot of ink. The title and URL of the webpage were included in the header of each printed page. The page number and total number of printed pages were in the footer of each printed page. Web pages consisting of multiple printed pages were stapled together in the upper left margin of the printed pages. Long web pages (i.e., in excess of 10 printed pages) had middle pages (mostly with repetitious content and/or formatting) excluded from the printing. As genre is characterized by specific types of content and format, these omitted pages should not have materially impacted the subjects' assessments.

Some pages from the original 133 were omitted from the final sample for various reasons. Some pages looked radically different in print (often because of the missing background). Some pages would not print properly. Pages of over-abundant page-types in the sample (newletters/articles, and top-level homepages for schools, for example) were removed to keep the size of the sample down, and to allow for the inclusion of page-types that have been shown to be relatively well-recognized (Rosso, 2002), and that were not well-represented in the sample, as categorized by the researcher, (forms, search engines, and

personal pages). Eight pages of these types were added to bring the final sample size of the printed web pages to 102.

A small, convenience sample of 3 people participated in the experiment. They were each given the printed web pages, and asked to separate them into like piles by genre. They were also asked to name the genres by writing the names on sticky notes and placing them on the piles. After the piles were complete, participants were asked to provide a short, one or two sentence, description of each genre, and then to describe the page characteristics that led them to place a page in that genre. Participants were also asked to identify the most and least representative pages in each pile, and to explain those choices. At any time during their explanations, they were allowed to move pages between piles, and to explain these moves. (See Appendix B for the actual instructions that were given and/or read to the subjects.)

Limitations

Before reporting the results and analysis of these experiments, it is important to consider the limitations of the methodology. First of all, the experience of viewing a web page on a screen, and reading a printed web page is very different. Although pages that were radically different without their backgrounds were removed from the sample, some differences did exist. For example, discerning what text was actually a link seemed to be harder. Viewing a page on a screen would be ideal. Unfortunately, sorting the pages in this format into easily accessible piles would be difficult. Other web page sorting studies have made this trade-off, including (Ryan, et al., 2002) and (Upchurch, et al., 2001). (Macskassy, et al. 1998) does not explicitly state whether web page sorting was done online or on paper, but the use of URLs in instructions to participants, and the small number of pages to be

sorted (16 or fewer) suggests that the web pages to be sorted may have been accessed online in that study.

Possibly because of the time-consuming nature of the task, subjects seemed not to read pages with blocks of text thoroughly. Different subjects focused on different aspects of the same page, and classified it differently. Also, there was not enough time to explore all the pages in the piles, to uncover possible inconsistencies in the subjects' sorting, in order to see how they would reconsider those pages.

Although the concept of genre was explained in writing and verbally, subjects seemed tempted to drift into subject-oriented classification, but they usually saw this during the exercise and corrected themselves. Subjects' unfamiliarity with the genre concept may still have hampered their performance.

Finally, although an attempt was made to eliminate pages that were nonsensical out of the context of their own websites, still, the context of surrounding pages was not there to help subjects judge the type of page. For example, a page with a title and a large paragraph of text was judged by one subject as a story, and another as an answer to a question (apparently because the title was in question format). In actuality, the page was a footnote explaining something on another page (to which, of course, these subjects had no access).

Results - Quantitative

Due to time constraints, not all requested information for all genres was obtained. The time spent on the exercise by the three individuals ranged from 1.75 to over 2.5 hours. In addition to the sorting results, subjects' comments were especially informative, and are presented and discussed in detail.

The number of genre piles made by each subject ranged from 24 to 28 for the 102 pages. The median number of pages per pile was 3 pages per pile, and the mode was 1 page per pile. The maximum number of pages per pile by subject ranged from 13 to 18 (see Table 4.1).

Table 4.1 Names of genres with greatest # of pages

Number of Pages	**Genre Name**	**Subject #**
18	Articles	1
16	Articles	3
15	Navigation	3
13	Story	2
12	Indicies/Table of Contents	1
12	Newsletters	2
11	Homepages	1
11	Links/Index	2

To determine the amount of agreement between subjects regarding genres, the overlap between pairs of genres between subjects was computed. Pairs of genres with non-trivial amounts of overlap are displayed in Table 4.2.

Table 4.2 Pairs of genre piles which have at least two pages in common, and in which the ratio of overlap relative to the total pages in the two piles is .25 or greater

Subject #s	# of pages in pile 1	# of pages in pile 2	# of pages in common between the two piles	Ratio of pages in common to all pages in the two piles	Genre Name	Genre Name
1 & 2	2	2	2	1.00	job listings or help wanted	job postings
1 & 3	3	3	3	1.00	poems	poems
2 & 3	5	6	5	0.83	forum	q&a forum
1 & 2	3	4	3	0.75	poems	poetry
2 & 3	4	3	3	0.75	poetry	poems
1 & 3	3	2	2	0.67	course descriptions	syllabus
1 & 3	18	16	12	0.55	articles	articles
1 & 2	7	5	4	0.50	diaries, weblogs or blogs	diary
2 & 3	3	3	2	0.50	database	search start
1 & 3	3	6	3	0.50	conversations, observations, or opinions	q&a forum
1 & 3	4	2	2	0.50	course listings	course lists
1 & 3	7	3	3	0.43	diaries, weblogs or blogs	diary
1 & 2	6	5	3	0.38	Q&A (not FAQ)	forum
2 & 3	11	15	7	0.37	links/index	navigation
1 & 3	12	15	7	0.35	indices/table of contents	navigation
1 & 2	3	9	3	0.33	course descriptions	course
2 & 3	5	3	2	0.33	diary	diary
1 & 3	8	4	3	0.33	resources or guides	advice
1 & 3	6	6	3	0.33	Q&A (not FAQ)	q&a forum
1 & 2	3	5	2	0.33	conversations, observations, or opinions	forum
2 & 3	13	16	7	0.32	story	articles
1 & 2	11	6	4	0.31	homepages	welcome pages
1 & 2	4	9	3	0.30	course listings	course
1 & 2	18	13	7	0.29	articles	story
1 & 3	2	7	2	0.29	program descriptions	blurb
1 & 3	6	3	2	0.29	Q&A (not FAQ)	faq/help
2 & 3	5	4	2	0.29	diary	personal website
2 & 3	4	5	2	0.29	card catalog	bibliography
1 & 2	12	11	5	0.28	indices/table of contents	links/index
1 & 2	3	12	3	0.25	publications, bulletins, newletters	newsletters
1 & 2	18	7	5	0.25	articles	reference

There were several pairings in which only one document overlapped, but that overlap constituted a large percentage of the total number of pages. These pairs are shown in Table 4.3. Conclusions about the recognizability of these genres must be tentative until more data on these types of pages are collected.

Table 4.3 Pairs of genre piles which have only one page in common, and in which the ratio of overlap relative to the total pages in the two piles is .5 or greater

Subject #s	# of pages in pile 1	# of pages in pile 2	Ratio of pages in common to all pages in the two piles	Genre Name	Genre Name
1 & 2	1	1	1.00	photograph	photo
2 & 3	1	1	1.00	photo	picture
1 & 3	1	1	1.00	photograph	picture
2 & 3	1	1	1.00	placeholder	web meta
1 & 3	1	1	1.00	definitions	fragment
1 & 2	1	1	1.00	email	email
2 & 3	1	2	.50	email	form
1 & 3	1	2	.50	email	form
1 & 3	1	2	.50	registration	form
1 & 2	1	2	.50	registration	contact form
2 & 3	2	1	.50	shopping	product for sale
1 & 3	2	1	.50	ads	product for sale
2 & 3	1	2	.50	template	fragment
2 & 3	2	1	.50	job postings	job ad
1 & 3	2	1	.50	job listings or help wanted	job ad
1 & 2	2	1	.50	biography	resume

Results – Qualitative

In addition to the descriptive data just presented, the dialogues with each subject regarding genre names, definitions, characteristics, and identification of most and least representative members of piles, provided rich insight into their (often internally inconsistent) categorization behavior.

Subject 1 remarked that one page (an article) was a “piece” rather than a “whole”. I believe that she was having difficulty distinguishing whether a page was complete unto itself, or was a part of a collection (which she placed in different piles). She also remarked that a page “might be further down in the site rather than at the top,” indicating that she was aware of the levels in a website hierarchy. There is no evidence that the knowledge impacted her sorting behavior (see Subject 2’s remarks below).

Subject 1 initially sorted “Indices” and “Tables of Contents” into separate piles, and when questioned, she maintained that they were different. However, in attempting to articulate the difference, she decided to combine the piles when she could not identify a clear distinction. Interestingly, her definition of “Homepage” includes what she calls a “mission statement” and a “table of contents”.

In examining her “Articles” pile, it seems that Subject 1 did not distinguish between popular and scholarly writing as a sorting criterion. However, she did seem to distinguish between spontaneous and edited (or polished) writing, as conversational pieces were separated from more formal writing, which included the text of a speech (which she stated she knew was a speech). Thus, a distinction between speaking and writing did not seem to affect her sorting.

Subject 1 mentioned that she did distinguish facts from narrative when explaining the differences between her “Biographies” and “Diaries/weblogs” piles. She did not group forms together, but separated them into singleton piles according to purpose: email, registration, search engine, etc. When the experimenter suggested the word “form”, she said she liked that as an overall term.

Subject 1 seemed to use words from the pages in her pile names, e.g., index, guide, newsletter, etc. However, one pile named “Q&A (not FAQ)” included a page with FAQ in its title. Due to time limitations, investigation into that pile was not undertaken.

Subject 2 noted that the relative amounts of pictures, links and text on a page influenced his classification. For example, his “Welcome Page”, which he described as a starting page even if it was at a middle level, had lots of pictures and minimal text. When Subject 2 was asked to explain the difference between his “Welcome Page” and “Links/Index” piles, he said that the latter had fewer pictures and less text.

Subject 2 generally placed pages with lots of text into “Reference”, “History” or “Story”. “History” involved the past, whereas “Story” tended to be shorter than “Reference”. Thus, text-length and situatedness in time seemed to be factors in Subject 2’s decisions. Interestingly, when these three piles together are compared with the “Articles” pile of either Subject 1 or Subject 3, there is over 50% overlap. It seems that Subject 2 may have sorted textual pages at a finer level of granularity than the other two subjects.

Like Subject 1, Subject 2 made a distinction between factual and narrative personal pages with “Resume” and “Diary” genres. When asked, Subject 2 admitted he might have gotten the name for the “Resume” pile from one of the experimenter’s examples given in the introductory explanation of the concept of genre. Also like Subject 1, Subject 2 classified forms separately according to purpose. However, unlike Subject 1, Subject 2 did volunteer the actual term, “form”.

Subject 2 also made a distinction between dynamic and static question answering with “Forum” and “FAQ” genres. “Forum” was described as one or more question and answer pairs (like an interactive discussion archive). FAQ was seen as a special type of

"Links/Index" in which questions are links to answers. Subject 2 also had a "Help" genre containing one page. "Help" was described like a "FAQ", except that the links were topics instead of questions.

Subject 3's "Q&A Forum" and "FAQ/Help" were very similar to those of Subject 2. "Q&A Forum" had 83% overlap with Subject 2's "Forum". "FAQ/Help" had 67% overlap with a combination of Subject 2's "FAQ" and "Help". Once again, it seems that Subject 2 created genres at a finer level of granularity. Also, Subject 3 had non-trivial overlap with Subject 1's "Q&A (not FAQ)": 33% with "Q&A Forum", and 29% with "FAQ/Help".

Subject 3 also stated that authorship made the difference between "Biography" and "Personal Website", the latter being autobiographical in nature. Unlike Subjects 1 & 2, Subject 3 grouped forms (except search engines) into "Form". Subject 3 stated that his textual genres of "Articles" and "Enews" were different in that "Articles" were not as time-sensitive, and were more topic-focused, rather than event-focused. Subject 2's more time-sensitive "Story" had an overlap with Subject 3's "Enews" of .21.

Subject 3 had four different genres for links-based pages: "Bibliography", "Navigation", "Full-text Index" and "News Index". "Bibliography" had links to broadly related resources, primarily at other sites. "Navigation" points to sub-categories (sub-pages) on the same site. "Full-text Index" points to the full-text of several works, or to the pieces of a single work. "News Index" points to "Enews" articles.

Discussion

Following is a summary of the web page aspects that seemed to influence subjects' perception of web page genre:

- whether the page is perceived to stand by itself as a logical whole, or whether it is a piece of a larger unit or collection
- the relative amounts of text, pictures and links on the page
- the perceived time-sensitivity of the information on the page (e.g., existence of a publication date, or column formatting indicating a "newsletter")
- genre-like words in the page title like "index" or "guide"
- length of text
- descriptive versus narrative text
- spontaneous versus edited text
- biographical versus autobiographical text
- lists of links
- perceived targets of links
- message header information like "To:", "From", "Re:", etc.

Differences in subjects' sorting behavior included using genres of varying generality, e.g., "FAQ/Help" versus the individual "FAQ" and "Help". Also, subjects seemed to classify by different aspects of the same page. For example, regarding the page in the Figure 4.2, Subject 1 seemed to focus on the page as a starting point, calling it a "Homepage", as opposed to her "Index/Table of Contents" genre. However, Subject 2 seemed to see the links as primary and classified it as "Links/Index", while Subject 3 seemed to pick up on the personal nature of the authorship, and called it a "Personal Website".

Figure 4.2 Sample page, which subjects classified by several different aspects

The same instructions were also given to an additional two-person team of subjects, but with a reduced set of 72 of original printed web pages. This was done in order to further explore the genres of the pages that were less consistently classified by the original three subjects. Even with 30 fewer pages to sort and explain, this pair of subjects took 55 minutes

to complete the entire exercise. However, many of the categories that they created were deemed to be idiosyncratic (e.g., “bullshit”), and their results were discarded. In the original rationale for this procedure, it was thought that having two participants talking over the decisions to come to consensus would be a useful alternative way of having single participants think aloud, to gather information about their decisions. Despite the lack of results from this endeavor, it was decided that enough usable genre information had already been collected in order to proceed with this next phase of the research.

User-based Refinement of Terminology into a Tentative Genre Palette

Given the same 102 printed web pages, participants in this next phase of research classified each page into one of 48 genres defined by participants in the previous study. The objective was to discover the amount of agreement between participants in using these genres to classify web pages from the edu Internet domain, and to discover which genres were preferred among those with similar names and/or definitions. Insights derived from participants’ sorting behavior were used to inform the creation of a genre palette to be used in subsequent research.

Although the overall goal of the palette is to provide genres that web searchers can include or exclude from their searches to improve the precision of their search results, the immediate goal of this study was to produce a palette of genres that users could assign to pages with a high rate of agreement among users. This agreement would then imply that the genres were recognizable by the users, and that the genres would be understandable to those using a search engine that included this capability.

Pre-Study Genre Palette Creation

In order for users to demonstrate high rates of agreement in assigning genres to pages, the genres needed to be conceptually mutually-exclusive, as study participants were asked to assign only one genre per page. Genres with conceptual overlap would have lowered the agreement rates. Thus, one aim of the palette creation process was to produce mutually exclusive genres.

At the same time, it is assumed that a genre palette cannot be both exhaustive and useful for searching. Focusing on the edu domain is one way to minimize this problem. However, within the bounds of mutual-exclusivity and the membership in the edu domain, it seems advantageous to have individual genres include as many pages as possible in order to keep the palette down to a manageable size for the users. Thus, the palette creation process balanced both needs: maximum coverage of specific genres, and the need for the genres in the palette to be mutually exclusive. Once the genres were found to be recognized by users in this study, subsequent research assessed the genres' ability to improve the interpretation of search results annotated by genre.

Before the commencement of this study, a list of forty-eight genre names and definitions was compiled using the data from the previous study (see Appendix C). A goal of the list compilation process was to leave the text of the genre name and definition pairs as close to those generated in the first study as possible. The rationale for this is that genres, if expressed in user-generated terminology, should theoretically be more easily recognized by members of the genres' discourse community. In the compilation, the original data was modified in the following ways:

- Genre names for which no definitions were obtained were excluded from the list.

- Genres which described concepts seemingly idiosyncratic and used by a single participant (e.g., “fragments”) were excluded from the list.
- Genres which described relatively uncommon types of web pages (e.g., “obituaries”) were excluded from the list.
- Plural genre names were singularized (e.g., “biographies” became “biography”) unless the name was a combination of separate document types (e.g. “publications, bulletins, newsletters”).
- Extremely similar genre name/definition pairs were combined into one. For example, “photo – no words” and “picture – page just has a picture on it” were combined into “picture/photo – page primarily contains a picture with few words on it”. Four genres (“email”, “job listing”, “picture/photo” and “poetry”) of the forty-eight genres were a result of this type of combination.
- Similar genre name/definition pairs that primarily overlapped but also contained slightly different concepts were combined into one name and definition, using more extensive wording alterations. For example, “diaries, weblogs or blogs - more like a narrative rather than highlights of a lifetime (i.e., not a biography)”, “diary - about me (or us) and what I've (we've) done in a story or narrative format” and “diary - time log of somebody's activities” were combined into “diaries, weblogs or blogs - a personal narrative or time log of activities (not a biography)”. Five genres (“biography”, “ FAQ”, “diaries, weblogs or blogs”, “forum/interactive discussion archive” and “indices/table of contents/links”) of the forty-eight genres were a result of this type of combination.

- Genres from different participants with the same names but different definitions (e.g. "article") were distinguished by adding a number to the name (e.g., "article –1" and "article – 2").

Procedure

Each study participant was given this list of genre name/definition pairs, the stack of 102 printed web pages (arranged in a different random order for each participant), an instruction sheet (see Appendix D), and a data collection form on which he/she recorded a genre for each web page. For each of the 102 web pages, the participant was given the option to either write a number from the list corresponding to a genre/definition pair which best described the page; or to provide his/her own suggestion for a genre name and definition, if none of those in the list seemed adequate. The participants were drawn from a convenience sample of approximately 10 college graduates, solicited through conversations with the experimenter's acquaintances, friends, and colleagues. Data collected included each of the ten participants' choices for the 102 pages, and any alternate genre names and definitions suggested.

Results

Participants took from 65 to 120 minutes each to complete the task, with an average of 90 minutes over all. Participants seemed fairly similar regarding the average agreement they had with the other participants, as measured by the number of other participants they agreed with on each page. With a possible range of agreement from 0 to 9 other participants, individual participants ranged from 3.4 to 4.2 participants agreed with, on average, per page.

Thus, there was no participant with a low (relative to the group) agreement level that might signal an individual with an atypical perspective on genres.

Of the 1020 genre decisions solicited from participants, 72 consisted of more than one genre, i.e., a participant listed more than one genre for his/her decision. Eleven of these 72 decisions were comprised of a single genre from the palette, plus a genre suggested by the participant. Three-quarters of the 72 multi-genre decisions were made by just two of the ten participants. (For statistical calculations, all decisions counted as one vote. Thus, if a participant assigned two genres to a page, each genre received half of a vote.)

Table 4.4 shows a frequency distribution of the maximum number of votes for a single genre that each page received. For example, the table shows that 9 pages had 9 or more votes but less than 10, comprising 8.8% of the pages. Twenty-five of the pages received 8 or more votes for a single genre. Thus, despite having 48 genres in the palette, a significant number of the pages were assigned to a single genre by a majority of the participants. Half or more of the participants agreed on a single genre for a page 60% of the time.

Table 4.4 Frequency distribution of the maximum number of votes in a single genre for each page

# of Votes	# of Pages	% of Pages (n=102)	Cumulative Values	Cumulative % (rounded)
2 – 2.9	4	3.9%	>=2	100%
3 – 3.9	18	17.6%	>=3	96%
4 – 4.9	19	18.6%	>=4	78%
5 – 5.9	17	16.7%	>=5	60%
6 – 6.9	12	11.8%	>=6	43%
7 – 7.9	7	6.9%	>=7	31%
8 – 8.9	12	11.8%	>=8	25%
9 – 9.9	9	8.8%	>=9	13%
10	4	3.9%	>=10	4%
Total	102			

Note: at least two people agreed on every page.

The flip side of that is that less than half of the participants agreed on 40% of the pages. Given the number of genres in the palette, and the overlapping quality of many of the name/definition pairs, this result is not totally surprising.

Table 4.5 shows the 25 genres for which there were pages that received 6 or more votes for a single genre (one possible indicator of the most recognizable genres).

Table 4.5 Genres with pages that received 6 or more votes

Votes per page	Genre Codes (see Appendix C for names and definitions)																										Total Pages
	A1	A3	A5	B2	C1	C3	C5	C6	D2	E2	F1	F3	F4	H3	I1	J1	N2	P1	P2	P3	P4	R4	S3	W1	XX		
10											1				2				1								4
9 – 9.9									1						1	2		1		3			1				9
8 – 8.9		1	1	1				2				3		1	1						2						12
7 – 7.9		1					1		1					1	1							2					7
6 – 6.9	1	1			1	1				1		1	1				1		1					2	1		12
Total Pages	1	3	1	1	1	1	1	2	2	1	1	4	1	2	5	2	1	1	2	3	2	2	1	2	1		44

Table 4.6 shows the remaining 24 genres of generally low agreement for which no page received 6 or more votes. (Note: the range 5 – 5.9 was chosen for completeness and consistency with Table 4.5. Actually, no page in Table 4.6 received more than 5 votes).

Table 4.6 Genres for which maximum agreement < 6 votes

Max # votes /page	Genre Codes (see Appendix C for names and definitions)																								Tot.
	A2	A4	B1	B3	C2	C4	D1	E1	F2	H1	H2	I2	J2	N1	N3	P5	P6	R1	R2	R3	S1	S2	S4	S5	
5-5.9		X					X	X		X		X	X			X					X		X		9
4-4.9			X		X				X		X													X	5
3-3.9	X																								1
2-2.9				X		X								X	X		X	X		X		X			8
1-1.9																			X						1

Table 4.7 Suggested alternate genre names and definitions by page number

Page #	Alternate Genre Names and Definitions
9	resources - further links to works by different authors related to the subject
17	course catalog - list of courses available by course number
34	events page
47	homework problems - list of tasks given as homework or for individual study test/quiz
52	curriculum description
57	links to subject examples - often found on programming sites
59	quotes - list of quotes
112	report - complete transcript of report, complete with bibliography
118	online exhibit
127	report - complete transcript of report, complete with bibliography
129	login page
132	handbook page
137	document - original source
140	case history – example illustrating the use of something
143	book description - similar to descriptions found on leafs of book covers
144	handbook section - section from employee handbook
146	reprint of author's work
149	resources - further links to works by different authors related to the subject
150	meeting report
161	reprint of author's work
162	document
167	search index - result of a search, giving suggested live links to further resources
170	redirection informational - "stop", "go back" under construction or error page placeholder for pages under construction housekeeping - programming info under construction update - information about a temporary situation or change
193	document (journal)
200	question

Appendices E through L show all the votes for all pages, with one table for each grouping of genres. The groupings subjectively appear to be somewhat related but no statistical tests have been applied to confirm or invalidate this. At this point, the genre groupings are used as a convenient way to explore the data in smaller chunks. Appendix M shows a grouping of genres that appear unrelated. (Note that page numbers are consistent

across these tables, such that if one wished to determine all votes for any given single pages, it would be possible, albeit inconvenient.)

Table 4.7 shows the names (and definitions when obtained) of the alternate genres for the pages for which they were suggested.

Overall, 31 genres were suggested for 25 of the 102 pages. Only twice did a page receive suggestions from more than one participant. However, it appears that these may be the most instructive cases. For example, page 170 received alternate names, from six of the ten respondents. This makes sense, considering that none of the categories suggested for page 170 in the previous study were included in this study's palette. Many of the alternative names expressed here represented similar concepts, suggesting that something in the vein of "under construction" may be a well-recognized genre. Unfortunately, it is probably not a genre that would be helpful for use in searching. For page 47, the three participants in the previous study judged it to be "Definition", "Fragment" or "Instructional". Of the three genres, only "Instructional" was included in this study, and it received 5 votes for page 47 (in addition to the three alternate suggestions). The suggestions (homework, test/quiz, problems) prompted this researcher to include an "Other Instructional Materials" genre in the palette for the next study (see chapter 5). Overall, there did not seem to be a pattern between pages receiving a suggested genre name, and the dominant genre voted for the page by the rest of the respondents. Each suggested genre was idiosyncratic to each participant. Duplicates in Table 4.7 were always the result of a single respondent suggesting the same genre for more than one page.

Discussion

Given that participants chose genres from a list of 48, many of which were extremely similar in nature, it would seem that the resulting agreement (half or more of the participants agreeing on one genre for a given page in 60% of the instances) is quite acceptable. However, analysis of what might have happened with the other 40% could yield useful insights for the development of subsequent genre palettes.

Explanations for participants' lack of agreement seem to be based on either the characteristics of specific pages and/or the nature of web genres in general. Some pages are comprised of multiple aspects. This phenomenon has been noted previously in the literature (e.g., Haas & Grams, 2000). An example of this in the current study is page 22 (see Figure 4.3). The page is one level below the school's main admissions page. It contains pointers to other parts of the admissions site (in addition to forms and a registration page), thus serving as a pointer to other pages. Four of the participants categorized the page in genres of the Pointers grouping (two chose "indices/table of contents/links"; two chose "welcome page"). However, there is also textual material about the admissions process and specific instructions related to the task of applying for admission to the school. Three participants categorized this page as "instructional" while two put it in the "about" genre, for a total of five in the Text grouping. Finally, the tenth participant called it a "registration form". Pages with a mélange of aspects simply cause generic disagreement among participants. A possible solution is categorization by the "primary" aspect. But how does one decide that? Possibilities include by proximity to the top of the page, by amount of square inches on the page, by the size of the aspect's font, heading or caption, or something else.

Figure 4.3 Page 22

It seems more reasonable, and has been suggested by others, e.g. Karlgren, et al. (1998), to

allow pages to belong to more than one genre. In fact, several participants did that by picking two or three genres from the palette for some pages.

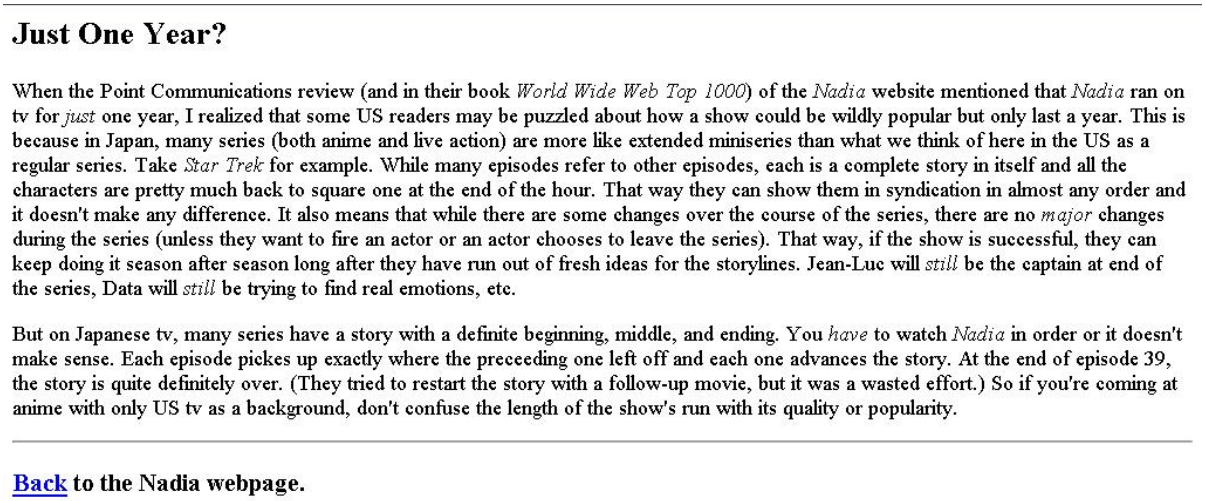
Another possible explanation for low agreement among participants is that sometimes a page's context is difficult to determine when the page is presented in isolation (as they were in this study). Not all pages are designed to be entry points. Page 128 is an example (see Figure 4.4).[5] The ten participants chose 6 very different genres in which to place this page containing a question as a heading, two paragraphs, and a navigational link at the bottom. Four chose "article-1". Two chose "conversations, observations, or opinions", and "FAQ". "Forum/interactive discussion archive", "history", and "review" were each chosen once. The page is actually a footnote from a page about an old Japanese cartoon series. With that knowledge, an "article" or "homepage" genre may well be most appropriate. In any event, pages that didn't provide enough clues as to their context were problems for this study. This, by itself, may not be an issue for document searching of machine-categorized web pages (especially since search engines may place users anywhere in a site) but it is an impediment for measuring the degree of recognizability of web genres.

A recognizable context, however, did not assure that a page's genre would be widely agreed upon by participants. Some pages, whose purposes seemed discernable, just didn't fit well into any genre in the palette. Page 132 is an example (see Figure 4.5).[6] It looks like a page from a school's "Career Development Center" for their "Pathways Program", judging from the heading and the table of links in the upper right of the page. It is an example of a "thank you letter", presumably to be sent to a mentor for counseling a student participating in

[5] The reader may wish to decide what genre that page 128 should be placed it (from Appendix C) before reading further.
[6] Again, the reader may want to venture a guess as to how page 132 would be categorized before proceeding further.

the program. The participants' ten votes were distributed as follows: "instructional" (4), "advice" (3), "welcome page" (1), "help" (1), "FAQ" (.5) and "handbook page", a suggested genre (.5). A "genre-less" page is not in itself an impediment to the idea of document searching by machine-categorized web pages. Some documents do not have typical purposes, forms, etc. Thus, the palette is not meant to be exhaustive. However, this situation makes measuring participants' genre agreement a little more challenging when pages are presented that, in the participants' perceptions, don't fit in the palette.

Figure 4.4 Page 128



Just One Year?

When the Point Communications review (and in their book *World Wide Web Top 1000*) of the *Nadia* website mentioned that *Nadia* ran on tv for *just* one year, I realized that some US readers may be puzzled about how a show could be wildly popular but only last a year. This is because in Japan, many series (both anime and live action) are more like extended miniseries than what we think of here in the US as a regular series. Take *Star Trek* for example. While many episodes refer to other episodes, each is a complete story in itself and all the characters are pretty much back to square one at the end of the hour. That way they can show them in syndication in almost any order and it doesn't make any difference. It also means that while there are some changes over the course of the series, there are no *major* changes during the series (unless they want to fire an actor or an actor chooses to leave the series). That way, if the show is successful, they can keep doing it season after season long after they have run out of fresh ideas for the storylines. Jean-Luc will *still* be the captain at end of the series, Data will *still* be trying to find real emotions, etc.

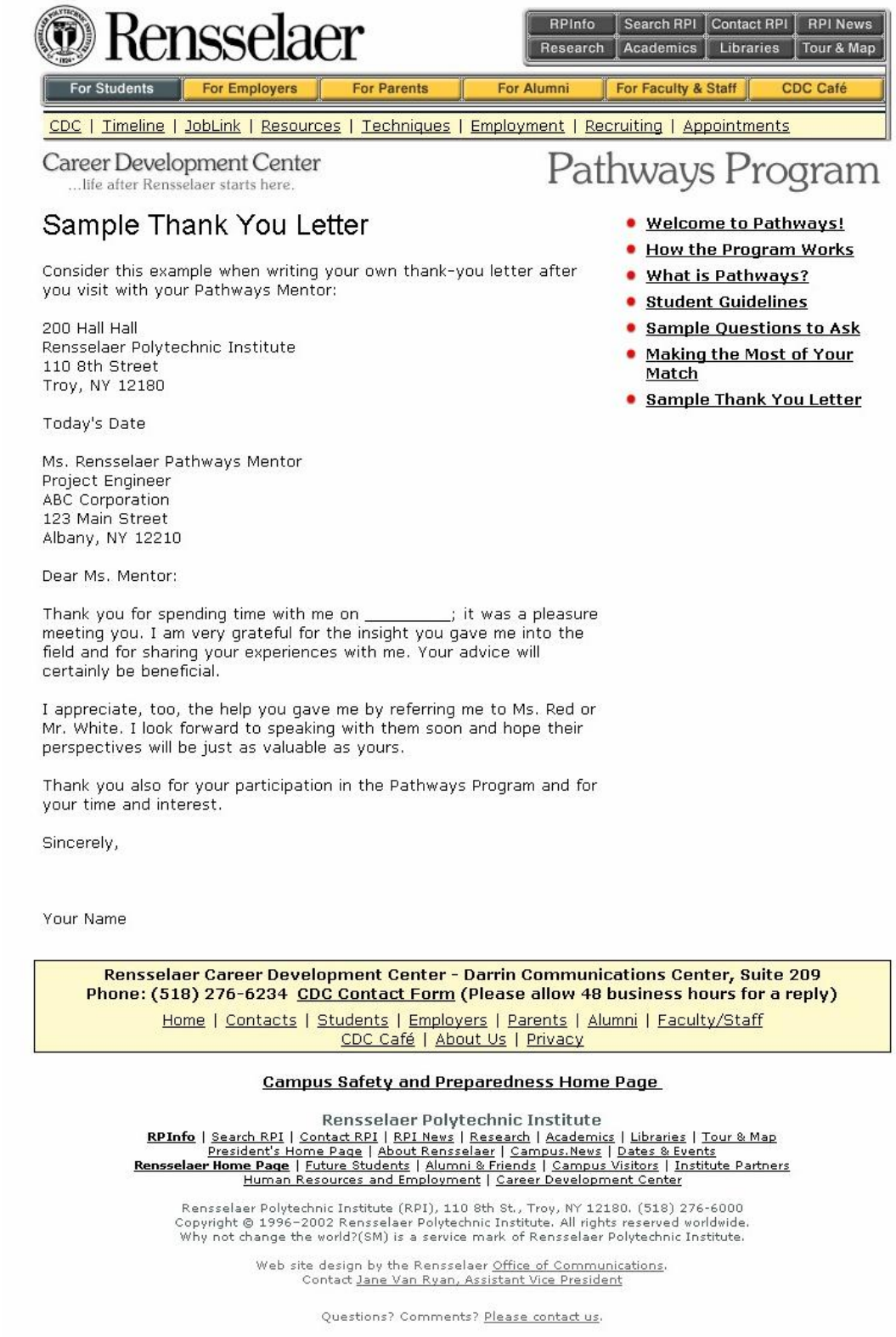
But on Japanese tv, many series have a story with a definite beginning, middle, and ending. You *have* to watch *Nadia* in order or it doesn't make sense. Each episode pickes up exactly where the preceeding one left off and each one advances the story. At the end of episode 39, the story is quite definitely over. (They tried to restart the story with a follow-up movie, but it was a wasted effort.) So if you're coming at anime with only US tv as a background, don't confuse the length of the show's run with its quality or popularity.

Back to the Nadia webpage.

Finally, there will typically always be some disagreement about genre because of genres' "fuzzy" nature. This is especially true of web pages.

> They are self-reliant and stand on their own but do not necessarily correspond to an entire document. Reading such pages is not always sufficient to understand and take in the document of which they are part or to index it correctly, i.e., to answer the following questions: What is this document about? For which user, for what purpose? (Prime-Claverie, et al. (2004), p. 1284)

Figure 4.5 Page 132



Rensselaer

RPInfo | Search RPI | Contact RPI | RPI News
Research | Academics | Libraries | Tour & Map

For Students | For Employers | For Parents | For Alumni | For Faculty & Staff | CDC Café

CDC | Timeline | JobLink | Resources | Techniques | Employment | Recruiting | Appointments

Career Development Center
...life after Rensselaer starts here.

Pathways Program

Sample Thank You Letter

- Welcome to Pathways!
- How the Program Works
- What is Pathways?
- Student Guidelines
- Sample Questions to Ask
- Making the Most of Your Match
- Sample Thank You Letter

Consider this example when writing your own thank-you letter after you visit with your Pathways Mentor:

200 Hall Hall
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180

Today's Date

Ms. Rensselaer Pathways Mentor
Project Engineer
ABC Corporation
123 Main Street
Albany, NY 12210

Dear Ms. Mentor:

Thank you for spending time with me on _____; it was a pleasure meeting you. I am very grateful for the insight you gave me into the field and for sharing your experiences with me. Your advice will certainly be beneficial.

I appreciate, too, the help you gave me by referring me to Ms. Red or Mr. White. I look forward to speaking with them soon and hope their perspectives will be just as valuable as yours.

Thank you also for your participation in the Pathways Program and for your time and interest.

Sincerely,

Your Name

Rensselaer Career Development Center - Darrin Communications Center, Suite 209
Phone: (518) 276-6234 CDC Contact Form (Please allow 48 business hours for a reply)

Home | Contacts | Students | Employers | Parents | Alumni | Faculty/Staff
CDC Café | About Us | Privacy

Campus Safety and Preparedness Home Page

Rensselaer Polytechnic Institute
RPInfo | Search RPI | Contact RPI | RPI News | Research | Academics | Libraries | Tour & Map
President's Home Page | About Rensselaer | Campus.News | Dates & Events
Rensselaer Home Page | Future Students | Alumni & Friends | Campus Visitors | Institute Partners
Human Resources and Employment | Career Development Center

Web site design by the Rensselaer Office of Communications.
Contact Jane Van Ryan, Assistant Vice President

Questions? Comments? Please contact us.

Writing on recognizing genre to aid evaluation of information quality of web documents, Sidler (2002) also comments that "the lack of form among web genres creates generic ambiguity" (p. 59). Recognizing a web page's genre is not always easy.

Many authors have noted that genres, especially web genres, are moving targets that evolve over time (e.g., Shepherd & Watters, 2004). For example, most "blogs" (according to Herring, et al., (2004)) are compatible with the definition of the "diaries, weblogs or blogs" genre in the current study. However, there are some pages called blogs with different purposes and/or forms. For example, some blogs function as support groups, or facilitate collaborative content creation (Krishnamurthy, 2002). Thus, the evolving nature of web genres may also be an impediment to genre recognition.

A possible solution to mitigate the effect of quickly evolving genres might be to consider using broader (relatively speaking) genres in the palette, with the hope that broader genres may evolve more slowly, and thus be easier to recognize.

Limitations

This study used genres expressed in user-generated terminology in hopes of clarifying the nature of user-preferred terminology for edu genres. Can users recognize genres as defined and described by other users? Is there shared knowledge among users regarding these genres, a knowledge that the definition of "genre" presupposes?

This study's genre palette had several limitations for answering these questions. First, the definitions used in the study were presented out of context. In the previous study, each genre definition was part of a participant's genre palette. In terms of a palette, each genre definition not only described a single genre, but also described what the rest of the genres in

the palette were not. That quality was lost in this study in which several palettes had been combined. Unlike a genre definition in a genre palette, each definition in this study had to stand on its own.

These genre definitions can also be considered to be out of context because the participants in the last study did not necessarily intend the definitions for a public audience. The participants verbally delivered the definitions to an experimenter who was giving them feedback as to his understanding. Thus, it is possible that a user may have possessed shared genre knowledge with another user, and yet not recognized it from that user's definition due to a breakdown in communication. That fact serves to justify the revision of a few definitions in the current study, although, at the expense of using "pure" user-derived terminology. For example, the definition of biography, given by a participant from the previous study, "a few pages about somebody" was changed to a "page primarily about a person" (which was intended to put less emphasis on the length of the page).

Finally, the use of the same sample of web pages in both this and the previous studies may have limited the results' generalizability to other web pages. In order to counteract this, the follow-up study, User Validation of the Genre Palette, (Chapter Five), presented participants with a different set of web pages to classify into a genre palette.

Creation of a New Genre Palette

Based on the results from the genre refinement phase, a genre palette was developed for the purpose of aiding the search process of edu web pages. The original list of 48 user-derived genres plus the newly acquired suggestions was trimmed down to a list of 18 (including a "none of the above" category).

Transforming the first list into the final list was not a straightforward, linear process, and is not easily documented by a chronological, step-by-step description. However, the general principles followed in the creation of the palette have been summarized, and their application to the list of 48 user-generated genres is illustrated.

1. Genres that dominated participants' responses for individual pages were kept. Dominant genres are those that are assigned to one or more pages with user agreement at 50% or more. Depending on the number of pages and the amount of agreement, this principle could be overridden by one of the others.
2. Genres that did not generally contribute significantly to user agreement (less than 50%) on the pages were eliminated, or combined with more dominant genres. These genres usually received few votes relative to others on the pages that were assigned to them. A genre might also be combined or eliminated if it seemed to pull votes from one or more, better-recognized genres. This situation is indicated by a genre receiving a small number of votes on pages which, as a group, were dominated by another genre.
3. Genres whose definitions contained overlapping concepts or similar concepts were combined or eliminated. Often, participants assigned these genres to many of the same pages. Multiple genres, all with agreement less than 50%, may be combined if they share multiple page assignments and some conceptual aspects, thus overriding principle #2. Application of this principle promotes the mutual exclusivity of the resulting genres in the palette.
4. After a tentative list of genres for the new palette was compiled, the genre definitions were edited for two reasons. First, sometimes it was deemed beneficial to increase the

generality of a genre in order to cover a wider variety of pages not covered by other genres in the palette. In these cases, care was taken to try to prevent the genre's concepts from overlapping with those of other existing genres. Wording was also changed in order to differentiate existing genres, genres which may seem to cover some of the same types of pages. The purpose of these wording changes was to cover as many pages as reasonably possible, without creating overlapping genres. (A workable solution for the web search problem may very well include overlapping genres; genres were intended to be mutually exclusive in this research to simplify the study.)

5. Because the palette was not intended to be exhaustive, a "none of the above" category was added in order to test participants' ability to recognize what pages were and were not covered by the palette.

The New Genre Palette

The new palette is shown in Appendix N. For ease of exposition, the changes to the original list of genres into the new palette are described by the previously mentioned genre groupings.

Academic Grouping (Appendix E). Genres C5 (course description) and S5 (syllabus) were combined because of their conceptual similarity, according to principle #3. All pages assigned to S5 were also assigned to C5, each by more users. Thus, participants seemed to prefer C5, and since the name was more general (principle #4), "course description" was kept. C6 (course list) was dominant for several pages, and pages to which C4 (course) was assigned overlapped completely with pages assigned to C6, so C6 was kept in accordance

with principle #1, and C4 was dropped because of low user agreement, and conceptual similarity, according to principles #2 & #3. P5 (program description) was also dropped because of low scores, in line with principle #2, and another genre in the new palette (welcome/homepage) could be re-worded to include program descriptions. This is one example of an application of principle #4. Finally, a new genre, "other instructional materials', was added because three participant-defined genre names (homework, problems, test/quiz) for one page suggested that this may be a well-recognized genre in the edu domain that was not covered by the palette. In addition, five participants used the I2 (instructional) genre to describe this page (although the genre's definition does not seem to apply). It was thus inferred that those participants' choices may have been based on the word, "instructional". Adding a new genre was an unusual action, not covered by the principles.

Personal Grouping (Appendix F). P1 (personal website) and D2 (diaries, weblogs or blogs) were kept because they were dominant on at least a few pages, according to principle #1. B2 (biography) was dropped because it detracted somewhat from P1 and D2, and biographies could be considered a sub-type of article. This demonstrates an application of both principles #2 and #3. R3 (resume) was dropped because it was not used much, and it also detracted from P1, (principle #2).

Question & Answer Grouping (Appendix G). C3 (conversations, observations or opinions) was dropped because it detracted from other genres. It was dominant on one page out of the 15 to which it was assigned. On the other 14 pages, user agreement was 3 or less (see Appendix G). This is an example of principle #2 overriding principle #1. F1 (FAQ) and H1 (help) were combined because of overlapping assignment to 4 pages, conceptual similarity, and the fact that neither one was that dominant on its own. In this case, principle

#3 overrode principle #2. F3 (forum/interactive discussion archive) was kept in accordance to principle #1, and was expected to be even more dominant after the elimination (mentioned above) of C3, which shared assignment with F3 on 7 pages (see Appendix G).

Text Grouping (Appendix H). A4 and A5 (article-1 and –2) were combined. There was significant overlap in their assignments to pages (13 pages assigned to both, out of a total of 23 pages), and several participants questioned the difference between the two genres. In examining the page assignments, no reasons for the difference between the genre assignments were discerned for the two genres. As one combined genre, A4 and A5 dominated assignments on 5 pages, which was more than any single genre. However, in 8 pages of the 23 total page assignments, 9 pages had user agreement for these combined genres of less than 4 votes. Generally, this much lower agreement might signal that this genre is detracting from other genres, and should be eliminated or combined with another, in line with principle #2. However, in those pages with low agreement in which a dominant genre was assigned, the dominant genre (advice, enews, review, or story) could be viewed as a sub-genre of article. Thus, in this case, principle #3 overrode #2, and as will be shown, the decisions to eliminate those other genres were made.

A3 (advice) was a fairly dominant category, with a majority of assignments on 4 pages, out of the 9 to which it was assigned. However, it was dropped due to conceptual similarity with FAQ/help, and also because it can be thought of as a sub-type of article, with which it overlapped in 2/3 of its page assignments. Principle #3 was applied to reduce conceptual overlap of the genres in the palette.

A2 (abstract), B3 (blurb) and H2 (history) all had low agreement and were dropped. I2 (instructional) had low scores and was dropped in favor of a new category (other

instructional materials), explained earlier. These decisions were all applications of principle #2.

Regarding the "news"-related genres, it seemed that participants didn't really have a clear conception of what news is. Even after combining two conceptually similar genres, N3 (newsletter) and P6 (publications, newsletters and bulletins), the highest score for a single page was 4, and that only happened once. Both were dropped (principle #2). E2 (enews) was assigned to 11 pages, 7 of which were also assigned to the article genres. In the two pages in which the E2 genre dominated, A4 (article) was the second-most assigned genre. Principle #3 was applied and E2 was dropped. The other "news" genre, N2 (news index), will be discussed with the Pointers grouping.

J1 (job listing) was dominant on the pages to which it was assigned, and was kept (principle #1). J2 (joke) was dropped because it was only assigned to two pages, and because it was thought that it might detract from pages that a user subjectively judged as humorous (principles #2 and #3). R1 (reference) was assigned to 12 pages, but never had more than 2.5 votes on a single page, so it was dropped (principle #2). Also, 7 of those pages were also assigned to one of the article genres (principle #3). R4 (review) was dominant on 2 of the 5 pages it was assigned to, and it pulled little away from other genres. However, it was dropped because of its multi-facetedness (reviews can be exposition, advertising, and/or summarization, for example), and because it can be considered as a sub-type of article (principle #3). S3 (speech) was kept because it was very dominant on one of three pages to which it was assigned, and also because of its lack of conceptual overlap with other genres in the palette (principle #1). S4 (story) was dropped (principles #2 and #3) because half of its page assignments overlapped with those of the article genres, and the other half of its page

assignments were to pages in which another genre predominated (e.g., S4 was assigned once to page 59 while J1 was assigned to it by 5 participants).

Pointers Grouping (Appendix I). I1 (index/table of contents/links) was kept because it was dominant on four pages (principle #1), and because it was the broadest of several genres which were assigned to many of the same pages (principle #3). N1 (navigation) was assigned to 22 pages, and never had a score higher than 2.83. It was dropped according to principle #2. H3 and W1 were combined due to high overlap (16 pages were assigned to both genres out of the 26 total pages assigned to either), and conceptual similarity evidenced by several participants' remarks (principle #3). Together, they dominated in 6 pages, and this was expected to increase with the elimination of N1 described above. B1 (bibliography) and C1 (card catalog) were dropped due to low scores (principle #2). The scores of F4 (full-text index) and N2 (news index) were not as low as B1 and C1, but F4 and N2 had high page-assignment overlap with other genres in the grouping, and it was hoped that I1, at a higher level of generality, would take up the conceptual slack (principle #3).

Forms, Search and Sales Groupings. Each of these groupings was collapsed into one genre (principle #2) because of extreme page assignment overlap (see Appendices J through L). Of the remaining genres (Appendix M), P2 (picture/photo) and P3 (poetry) were kept because they were very dominant (principle #1). A1 (abstract) was assigned to 8 pages, dominant on one page, and scoring 3 or less on the others. A1 was dropped with the expectation that other genres (e.g., article) could pick up the slack (principle #3).

Palette Refinement

According to principle #4, with a tentative set of genres for the new palette selected, changes were made to their definitions to improve the overall palette in terms of inclusiveness of pages, and mutual exclusivity of the genres (see Appendices C and N). In order to increase the generality of the category, the final definition for "article" is a pared-down (and hence, less restrictive) version of the definition of article-1. "Diary, weblog or blog" was made singular, and changed to imply that biographies belong in the "article" genre, and not here. "FAQ/help" is a simple combination of its constituent genres. "Form" was given a little more explanation, and differentiated from a search engine. "Index/ table of contents/links" was rewritten to include lists that don't consist of links (i.e., text items), and to differentiate it from the other genres whose pages might primarily contain links. "Personal website" has additions to differentiate it from a biographical article and a homepage. "Picture/photo" was expanded to include pages with lengthy captions. "Product for sale/shopping" was given the more general definition of the previous "product for sale", with additional text to differentiate it from a review article. "welcome/home page" contains the definition from the previous "welcome page" genre plus information to make it applicable to more than just organizations (as in the previous "homepage" definition).

This new genre palette was tested for user recognition with new pages drawn from the edu domain in the next phase of this research.

This page left intentionally blank.

CHAPTER 5

USER VALIDATION OF THE GENRE PALETTE

This work builds on the previous studies towards the goal of creating a user-recognizable genre palette. Following is a summary of the work so far.

(Also see Table 3.1.)

1. Three participants, individually, were given approximately 100 printed web pages, and asked to sort them into stacks, according to genre. Participants were to give each stack a name and a brief description.
2. A tentative genre palette of 49 genres was created, consisting of genre names and definitions collected in the previous step, with user terminology retained as much as deemed appropriate.
3. Ten participants, individually, were asked to assign a genre from the tentative palette to each of the same printed web pages used in step one above.
4. From the data collected in step three, an 18-genre palette (names and definitions) was created.

The objective of this part of the research was to show that the genres in this 18-genre palette can be recognized by users. Study participants were given the genre definition list, and asked to classify a new set of 55 web pages.[7]

[7] It has been shown that people can use written definitions of made-up categories ("ad hoc categories", such as these genres) to successfully place things into categories (Barsalou, 1982).

In order to measure the degree of user recognition of our genres, we did not rely on traditional coder-responder agreement. Instead, a user-based measure of consensus was employed, according to the concept of "user warrant". This concept was discussed by Lancaster (1986) to justify the inclusion of certain indexing terms in retrieval systems. Haas & Hert (2000) explained that user warrant "states the importance of looking at the community of users of the target information, and determining how they identify it." In this study, we apply user warrant as follows: if the majority of the participants agree that a web page is of a particular genre, then it is. The higher the level of agreement, the more we might infer that a particular genre is socially recognized.

Collection of Web Pages

A computer program was developed to download webpages, from which a 55-page subset was to be selected to be shown to participants to categorize. All genres in the palette were to be represented by at least two pages (as determined by the researcher), and pages of varying degrees of typicality of their genres were desired. Genres which seemed less "well-formed" (from the previous studies, (i.e., article, index/table of contents/links, welcome/homepage) were given more pages. A few pages which didn't seem to fit any of the genres well (again, as determined by the researcher) were included to see if participants would use the "None of the above" category for them.

The algorithm used to collect the pages was similar to that used in the collection of the 102 printed web pages that were shown to participants in an earlier phase. The pages were collected from the edu domain from search engine results from Google, produced by single-term queries, with each query consisting of one of the most frequently-used words in

the English language and limited to the edu domain. The idea is that a high frequency word (like “she”, for example) is not significantly related to the content or “kind” of page retrieved from the search engine. The list of high frequency words used was obtained from the Brown corpus (Francis & Kucera, 1982). (See Figure 5.1 for the algorithm for collecting the pages.)

(Although the algorithm was similar to the one used previously, different pages were retrieved since these were collected in August 2004, whereas the previous sample was taken in April 2004.)

Figure 5.1 Algorithm for collecting a pseudo-random sample of web pages

1. *Starting with the most frequently used word (Francis & Kucera, 1982), use the word as a single-term query to the search engine.*
2. *Download the pages (and associated files: images, etc.) that are the 200^{th,} 400th, 600th, and 800th results of the query.*
3. *Repeat Steps 1 & 2 with the next most frequently used word until the sample is complete.*

The computer program developed to do the downloading in this study tried to download 252 pages – 4 search results for 63 queries. However, the actual number of pages collected was affected by URLs whose pages could not be retrieved. Reasons other than unavailability for not including pages for consideration for the test set included: pages where all or many of the images were “broken”, pages that opened additional pop-up windows, pages where non-standard plug-ins were needed to view some of the content, pages in non-HTML formats, and pages with noticeably obscene words or sexually-oriented topics.

In order to collect enough pages to meet the criteria for the genre palette, the downloading program was run a total of three times. In the last two, the algorithm was modified to collect the 175th, 375th, 575th, and 775th, and the 150th, 350th, 550th, and 750th search results from each query, respectively.

Data Collection

Participants were originally recruited from the faculty, staff and students of a small, Southeastern private college. They were solicited through an announcement on the college's internal home page, and through a couple of email messages to all faculty and staff. Prospective participants were invited to become part of the study by accessing the survey through the study's URL. Seventeen people completed the entire task in a 3-week time span. Because a greater number of participants was desired, an email was sent to all faculty, students and staff (who hadn't opted out of receiving mass emailings) from a large, Southeastern public university. Two-hundred-forty additional participants from the larger institution completed the task over the next eleven days, for a total of 257 participants.

A website was developed to allow participants to view and categorize the web pages, one at a time. A small frame at the bottom of the screen allowed participants to assign each page to a category. A small frame at the top displayed each page's original URL. At all times, the participants were given access to a page with the genre names and definitions, which linked to multiple example pages of each genre in the palette.

Before the pages were viewed, some demographic information was collected from each participant. After that, five "practice" pages were then shown to each participant, always in the same order. The intent was to minimize any potential training effect that could influence the results. After the practice pages were viewed and categorized, the 50 test pages

were presented in a different, random order to each participant to minimize any potential order effects.

After all pages were viewed, participants were allowed to submit any feedback or comments they had about the task, and to volunteer to be contacted again in order to provide additional data for the study. (Screenshots of the initial page, the demographics/instructions page, a sample categorization page, and the feedback collection page are provided in Appendices O through R.)

Results

Participant Demographics

Table 5.1 Participant demographic information

Participants	Average Age	Average # of hours spent using the Web per week	Average # of hours spent searching the Web per week
141 students	23.3	19.9	8.0
116 non-	37.9	22.1	9.1
Average	29.9	20.9	8.5
Range	18-72	3-100	0-60
Median	26	15	5

The demographic information for the 257 respondents who completed the survey is shown in Table 5.1. Eighty-four of the 257 participants left feedback using the feedback collection page (see Appendix R). Fifty-nine of them left comments having to do with their experience of the categorization task, 17 left only contact information for possible follow-up, and 8 made extraneous comments such as: “You have a challenging task. Best of luck.”, “Are

you just taking advantage of students to gather search engine data?", and "Thanks!". A substantial body of comments, richly characterizing participants' experiences, is reported on later.

Participant Agreement

Consensus was the measure by which we determined whether a specific webpage was recognizable to the participants as belonging to a particular genre. Consensus was considered to be the largest percentage of participants that assigned a specific genre to a page (of all genres assigned to the page). Presentation of the data for the experimental and practice pages is combined because analysis of the data indicated no training effect. Participants' consensus for the practice pages ranged from 50% to 94%, with an average of 78.5%.

Figure 5.2 Consensus per page

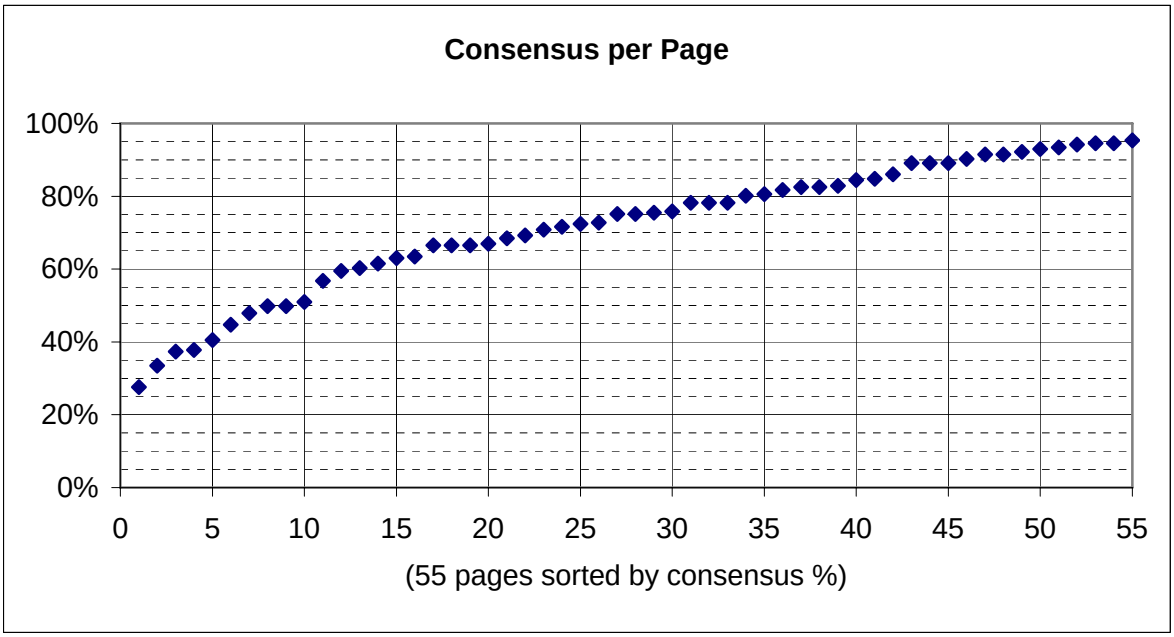


Figure 5.2 shows the consensus percentage for all 55 pages (50 experimental and 5 practice pages) sorted by consensus percentage. Forty-eight pages of the 55 achieved a consensus of 50% or higher (including all of the practice pages). (These pages will be referred to as the "consensus pages". The other seven will be called the "non-consensus" pages.) Twenty-three of the 55 achieved a consensus of 70% or greater. Average agreement per page, on all 55 pages, was 71.9%. Inter-participant agreement was 58.3%, with a Cohen's kappa of .55.

The choice of the 50% level for separating consensus pages from non-consensus pages was based on two considerations. First, it was thought that, given that there were 18 genres in the palette, any single genre garnering 50% agreement would, in most cases, be significantly higher in agreement than any other single genre for any single page. Also, given that the purpose of genre identification here is to improve search, it is conceivable that the use of a genre, only agreed upon by half the users, might still be helpful for enhancing some searches.

Two measures to assess users' recognition of web page genres were used. The first was the users' average agreement over the pages in which a consensus of 50% or higher was achieved for a particular genre (average %, see Table 5.2). For example, on each of 2 of the 55 pages, the "course description" genre was assigned by 50% or more of the participants. As another example with one page, 93% classified it as "poetry". On the other page, over 95% did so. Thereby, the average percentage of consensus was 94%. Note that, by definition, the lowest possible percentage is 50%. Thus, the possible range of the measure is 50-100%. The measure is potentially volatile because it is the average of a small number of pages. The

largest number of pages in one genre for which the participants reached consensus for a given genre, was 6.

Table 5.2 Participant consensus by genre for the 48 pages with consensus of 50% and higher

Genre	genre as Consensus Genre		other genre as Consensus Genre	
	# pages	avg %	# pages	avg %
article	6	67.1	42	1.6
course description	2	94.2	46	2.4
course list	3	76.3	45	0.3
diary, weblog or blog	3	85.2	45	1.1
FAQ/Help	2	84.8	46	1.0
form	3	69.6	45	1.0
forum/interactive discussion archive	3	74.7	45	1.0
index/table of contents/links	2	61.1	46	1.7
job listing	2	82.1	46	0.0
other instructional materials	3	75.1	45	2.6
personal website	4	76.1	44	1.4
picture/photo	2	88.3	46	0.2
poetry	2	80.0	46	0.4
product for sale/shopping	2	79.2	46	0.2
search start	3	84.0	45	1.3
speech	1	66.5	47	1.1
welcome/homepage	4	79.6	44	2.0
NONE OF THE ABOVE	1	60.3	47	5.2
Average by Genre	2.7	76.9	45.3	1.4
Average by Total Participant Decisions	-	76.8	-	23.2
# of Pages at 50%+ Consensus	48			
# of Pages at less then 50% Consensus	7			
Total Pages (including 5 practice pages)	55			

The second measure provides us with an indication of “false hits” for a genre, i.e., the percent of times that a genre was chosen as a page’s class, when some other genre received

consensus of 50% or higher for that page (again, see Table 5.2). Using the "course description" genre as an example again, 2.4% of the time that a participant classified a page as a course description, a genre other than the course description genre was the consensus genre of the overall group of participants.

These two measures, taken together, provide a more complete picture of the users' recognition than either measure alone. For example, even though the users achieved consensus on the "course description" pages with an average of 94.2% agreement, 2.4% of the time that users selected a page as a "course description", it was not the consensus genre. To state it another way, although "course description" pages (as identified by the users) received 94.2% consensus, 2.4% of the time a page was classified as course description, the choice was wrong (again, as determined by the consensus of the users' choices). Note that, according to Table 5.2, 76.8% of participant votes were in agreement on the consensus genre for each these 48 pages. Conversely, 23.2% were "wrong", according to participant consensus.

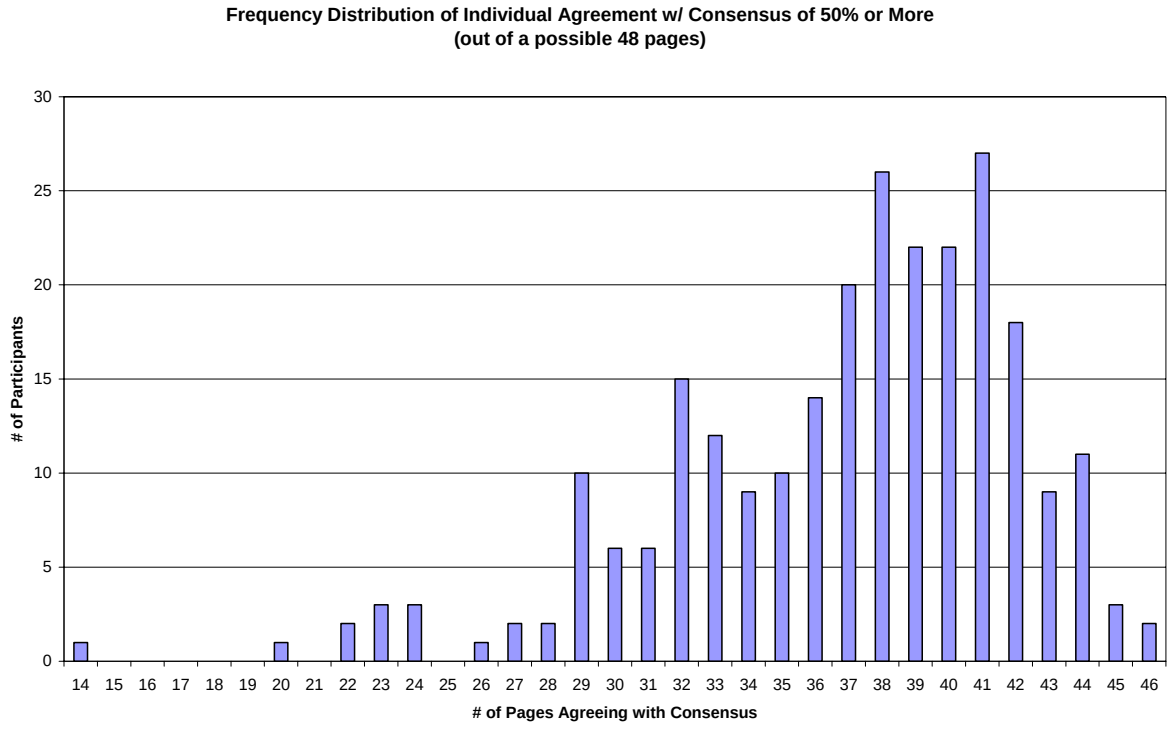
Table 5.2 shows that the genres vary in terms of recognition. However, it is not obvious how one might rank the genres on a recognizability continuum, given that is unclear how to balance the two measures of consensus and false hits. For example, "course description" had the highest consensus of all the genres at 94.2%. However, it had one of the worst false hit rates at 2.4%. Is course description more recognizable than "picture/photo" which had an 88.3% consensus (which was second best) and false hits of .2% (also second best)?

For the sake of analysis, if the averages of the two measures are used to roughly divide the genres into three groups of high, medium and low recognizability, it is clear that

"picture/photo", "job listing", "poetry", "product for sale/shopping", "FAQ/Help", "diary, weblog or blog" and "search start" are highly recognized. It is unclear whether "course description" and "course list" would be in the high or medium recognizability groups because of the disparateness of the two measures. Clear members of the medium recognizability group would be "personal website", "forum/interactive discussion archive", and "form". It is unclear whether "welcome page" and "speech" should be in the medium or low groups. Clear members of the low recognizability groups are "article", "index/table of contents/links", "other instructional materials", and "none of the above".

Participants varied in the extent of their agreement with each other on the 48 pages with a consensus of 50% or higher (see Figure 5.3). For example, one participant agreed with the consensus only 14 times out of 48, while 2 participants agreed with the consensus on 46 of the 48 pages.

Figure 5.3 Frequency distribution of individual agreement with consensus of 50% or more



Participant Feedback

The abundance of detailed comments that participants left on the feedback page (Appendix R) was unanticipated. Several made comments about the general ease of the task:

- “For most pages a first glance was all that was needed.”
- “Most of the categorizations just seemed obvious and instant to me.”
- “The catergorization task was pretty straight-foward.”
- “For the most part the selections were pretty clear.”
- “Most of the sites seemed terribly obvious……Again, this all seemed so obvious I'm wondering what you were looking for. Maybe less experienced users have trouble with this?”
- “With my web experience, usually one scroll down the page was enough to tell what the site was.”
- “Honestly, people who use the internet (as much as UNC Chapel Hill Students do) should have no trouble determining what the site is meant to do.”

Many (including some who made the previous comments) indicated that there were some pages that were problematic to categorize:

- “I had some trouble fitting sites into categories provided.”
- “The list of categories was daunting and difficult to navigate.”
- “Categorizing web pages, though, is extremely difficult because of the great variety of information out there. “
- “The only problem I had was that some did not fit categories neatly.”
- “As with all things, some web pages are quite hard to categorize. They don't fit into an established mold.”
- “Some pages are very confusing in the way they are designed and it is difficult to determine exactly what is the goal or intent of the various pages.”

Participants provided rich commentary on the page elements they used to make their categorization decisions. Most frequently mentioned aspects were keywords and phrases, URLs, and titles. Other aspects mentioned were colors, typeface, layout/style, backgrounds, layout/format/style, lack of graphics, presence of buttons, and generic images like the “shopping cart”. Following are some representative quotes:

- “If it wasn't immediately obvious, I would scan the URL - lots of semantic clues there (e.g., the word 'blog', etc).”
- “I identified speeches as such when they said, "given by" or something similar somewhere on the page.”

- "… large titles often made things clear, i.e. "So-and-so's Weblog" or "FAQ.""
- "Bold colors, creative looking, polished I felt the most comfortable with identifying the page quickly. Things only in normal type, black in white, it was a bit more difficult to tell."
- "Articles, speeches, text (etc) seem to have more of the bare background of either white or grey."
- "Whenever a site made reference to personal life issues, as in "my house," "my mother," "where I work," etc. that seemed a pretty good indicator of a personal site. Naturally when a price was listed or a credit card number requested, or the word "sale" appeared I figured this was a site designed to sell something. Obviously, any site that had boxes to fill in or check was a form."
- "Personal webpages normally have fun colors and use first person often. Course descritions list courses and describes them, or describes one course. FAQs normally are listed with bold print at top or see question/answer format."
- "Articles and Archived Forums were also easy to spot, since they typically do not involve graphics."

As noted in earlier quotes, participants sometimes reported that some pages didn't seem to fit in any of the genres. To remedy that, many genres were suggested to be added to the palette, e.g., survey, test, quiz, events schedule, search results, calendar, directory, abstract. Here are some comments on pages that did not seem to fit:

- "A syllabus is different from a course description and a list of course descriptions is slightly different from a simple listing of available courses, but maybe I am splitting hairs."
- "A Search Results category should be included. Some pages were the results of a search, and categorized as None of the Above."
- "There wasn't a category for calendars, which made some of the calendars go to different categories than others."
- "…the site which featured e-books that were free to read online but could also be bought in PDF form didn't seem to clearly fit into a category."
- "…the personality test used a form, so I classified it as a form, but I wouldn't say its primary /purpose/ was to function as a form. If "quiz" were an option, that would be more accurate.".
- "i didn't know whether an abstract for an article counted as an article or if it should be "none of the above". one page provided a link to the actual article, so i chose article."
- "There seemed like there should have been an seperate category for such things as lists of faculty, or departments. I categorized them as indexes for lack of a better category. Also I wasn't to comfortable putting article abstracts in any of the categories."
- "Didn't know how to rate personal information in the context of institutional setting."

At times, participants seemed to find more than one applicable genre for a page:

- “Some webpages seemed to fall into 2 categories, such as pages that are both course lists and course descriptions. Sometimes I couldn't tell if something was a personal webpage or a home page.”
- “I found it difficult to categorize pages which I felt fit multiple categories where both seem equally prominent, e.g. What is a home page with a search engine on it?”
- “Some things were very grey...the page I would describe as an article about poetry, for example...had a lot of poetry in it, but I still characterized it as an article.”
- “One site was just a picture but had a URL that marked it as clearly for a class. I think I put it under Other Instructional Materials.”

Some participants found some genres to be vague or ambiguous:

- “I categorized a lot of the pages as "other instructional materials." It might help to make this category more clear or to split it out into more well-defined subcategories.”
- “"Article" category seemed to be a bit ambigious. Usually one thinks of an "article" as referring to well-developed paragraphs about a topic. But the description here said "information about a topic", so would a webpage that, say, lists the properties of hydrogen, be considered an article?”
- “I found the welcome/homepage a bit disconcerting. Many pages seemed that they were welcome pages, but definately not homepages, wheras others were in fact homepages.”
- “It was hard to decide whether some pages were course lists or descriptions because they were lists with mini-desciptions.”

A few comments suggested that participants may understand typical characteristics of some genres:

- “Home pages are usually well-delineated with "About Us" or "Welcome" with sidebar lists. It is sometimes harder to decide between blogs, forums and personal websites.”
- “Usually one thinks of an "article" as referring to well-developed paragraphs about a topic.”
- “For example, some appeared to be FAQ/help even though they did not reflect the typical format of such sites.”
- “Articles and Archived Forums were also easy to spot, since they typically do not involve graphics.”
- “Online surveys are usually technically forms, but they seem to serve a different purpose.”

Some participants noted that, if they had been allowed to change some of their categorizations, they would have.

- “Early on in the study, I miscategorized a page. The webpage was an article written by a woman. I didn't realize that 'Article' was one of the options.”
- “Mislabled a few of the pages as blogs before I realized that there was a "forum/discussion" category.”
- “I realized later that there are two pages that should have been categorized differently.”
- “I found myself wanting to go back a couple of time to see what I had answered before or to change an answer. I am assuming the inability to go back was intentional but I forgot that there was an article option and wanted to go back and change once I realized it.”
- “Some of the pages I categorized as "article," "other instructional material" or "none of the above" were clearly lectures/speeches. I realized at the end that I was very inconsistent in how I listed them.”
- “I accidently clicked "article" instead of "speech" on one website that was a the text of a sermon.”
- “I miscatagorized a 'start search' as a I think 'none of the above'”
- “It may have been helpful if I could have gone back to pages and changed my selection because I became more comfortable "catagorizing" the web pages as I worked through your set.”

A couple of participants appeared to be questioning the purpose of identifying a page’s genre.

- “It was interesting deciding on some of these. The category of a page is hardly a consideration when "Where's the information?" is the purpose of the visit.”
- “Normally, I wouldn't seek to classify web pages in order to know whether they were relevant to my interests or objectives; either the information would interest me or not, continue to inform me or not, and I'd move on to the next search technique. “

Finally, not everyone may have understood what a blog is, as evidenced by one of their comments, “What’s a blog?”

In summary, the comments provided much insight into participants’ experiences of single-genre webpage categorization: problems with pages fitting multiple categories, problems with pages fitting no categories, and a general recognition of the characteristic formal elements of many of the genres. The comments also provided excellent feedback that could inform future research in terms of improved experimental procedures and increased

clarity of genre definitions and coverage of genres in the palette. More in-depth mining of this rich source of qualitative data could provide further insights in the future.

Failure Analysis

A failure analysis was conducted on the 7 pages which failed to receive a consensus of at least 50% for any of the genres (the "non-consensus" pages). For these low agreement pages, it is interesting to note that they comprise 12.7% of the sample, but received 28.1% of the "none of the above" genre votes. Following are short descriptions of these pages, the votes they received for each genre, and references to their images in the Appendix.

Page 31 (see Appendix S) looks like a homepage but describes a course, but not a typical course. It is a six-week intensive course for college graduates. The votes for page 31 are shown in Table 5.3.

Table 5.3 Votes for page 31

Genre	**# Votes**	**%**
course description	104	40.5%
welcome/homepage	51	19.8%
none of the above	38	14.8%
FAQ/Help	26	10.1%
article	14	5.4%
other instructional materials	9	3.5%
job listing	6	2.3%
index/table of contents/links	3	1.2%
course list	2	0.8%
form	2	0.8%
diary, weblog or blog	1	0.4%
picture/photo	1	0.4%
Total	257	100.0%

"Course description" received the most votes with 40.5%. It is unclear whether participants were conflicted over whether this was a "course description" or a

"welcome/homepage" or whether they thought it was both. The page is not typical of either genre, which might explain the 14.8% voting for "none of the above". The page contains a subheading in the form of a question, and contains a link to frequently asked questions, which could explain the "FAQ/Help" genre receiving 10.1% of the vote. Other than the 5.4% voting for article, it seems difficult determine why there are votes for other genres. A link to application forms may have prompted the 2 votes for "form". Although the page mentions being for recent college graduates who may have decided to make a career change, that doesn't seem like sufficient reason to put it in the "job listing" category, as 6 did. There is a photo, and that may have been enough for one to vote "picture/photo". The handful of obscure votes is typical for almost all of the 55 pages. Overall, the form and content of page 31 do not seem to come together in a typical generic format, which could explain the lack of participants' consensus for the page.

Table 5.4 Votes for page 69

Genre	**# Votes**	**%**
index/table of contents/links	97	37.7%
none of the above	65	25.3%
other instructional materials	36	14.0%
search start	28	10.9%
picture/photo	11	4.3%
welcome/homepage	11	4.3%
forum/ interactive discussion archive	4	1.6%
article	2	0.8%
form	2	0.8%
FAQ/Help	1	0.4%
Total	257	100.0%

The votes for page 69 are shown in Table 5.4. Page 69 (Appendix T) features a U.S. map which links to 50 weather data pages, one for each state. Below the map is a table with

text links to the states, which serves as an alternate way to link to the data. It may not be a typical page, but it seems like a typical application of an HTML image map. Thirty-seven-point-seven-percent thought it was an “index/table of contents/links” page. However, 25.3% thought it belonged in “none of the above”. “Search start” (10.9%) and “other instructional materials” (14%) were the other two significant vote-drawing genres. Given the image, “picture/photo” (4.3%) is an understandable choice. It is unclear why someone would describe it as an “article” (2 votes). “form” (2 votes) or “forum/interactive discussion” (4 votes). Unlike page 31, which had two genres (other than “none of the above”) with 20% or more, the pattern here seems to be one of confusion, with the most people seeing the links as the characteristic feature.

Table 5.5 Votes for page 122

Genre	**# Votes**	**%**
product for sale/shopping	86	33.5%
article	79	30.7%
other instructional materials	24	9.3%
index/table of contents/links	22	8.6%
search start	20	7.8%
none of the above	16	6.2%
welcome/homepage	5	1.9%
course description	1	0.4%
diary, weblog or blog	1	0.4%
FAQ/Help	1	0.4%
form	1	0.4%
personal website	1	0.4%
Total	257	100.0%

Table 5.5 shows the votes for page 122 (Appendix U). Page 122 seems to be a clear example of a page belonging to multiple genres. The page sells scholarly books online, and also reproduces the text online. Votes were split almost evenly between “product for sale/shopping” (33.5%) and “article” (30.7%). A link to the table of contents of the book may

explain the “index/tables of contents/links” genre receiving 8.6% of the votes. A search box for searching the book may explain “search start” receiving 7.8%. “Other instructional materials” received 9.3%, and “none of the above” received only 6.2%. It would seem that the allowance of multiple genres per page could result in more consensus for this page.

Table 5.6 shows the votes for page 154 (Appendix V). Page 154 was just 5 votes short of a consensus for the “other instructional materials” genre (47.9%). Page 154 is an answer key for a quiz on facts about alcohol consumption, produced by an academic center for alcohol studies. The page doesn’t seem to be related to any course, and may not be considered instructional by some people. This interpretation is consistent with the fact that the next two highest vote-getting genres at 14% each were “FAQ/Help” and “none of the above”. “Form” received 10.1%. The only other genre to get a significant percentage was “article” at 8.9%, possibly due to the lengthy answers for some of the questions. It would seem that a likely reason for the lack of consensus for this page is that it doesn’t quite fit into any of the genres in the palette.

Table 5.6 Votes for Page 154

Genre	**# Votes**	**%**
other instructional materials	123	47.9%
FAQ/Help	36	14.0%
none of the above	36	14.0%
form	26	10.1%
article	23	8.9%
forum/ interactive discussion archive	5	1.9%
personal website	3	1.2%
welcome/homepage	2	0.8%
index/table of contents/links	1	0.4%
job listing	1	0.4%
product for sale/shopping	1	0.4%
Total	257	100.0%

Table 5.7 shows the votes for page 158 (Appendix W). It is a four-year plan for helping undergraduates to decide on their eventual career direction, sponsored by a school's career center. This page did not seem to fit any of the genres, garnering the highest amount of "none of the above" votes of any page (37.4%). It is not readily apparent why 30.7% put this page in "other instructional materials", given that this page does not seem to be course-related. The next highest genre in terms of votes was "FAQ/Help" (11.3%). All other votes were spread across nine genres with no more than 5.8% for any single genre. In all, this page received votes in 14 of the 18 genres (as compared to 8.5, the average number of genres that received votes for a single page).

Table 5.7 Votes for page 158

Genre	**# Votes**	**%**
none of the above	96	37.4%
other instructional materials	79	30.7%
FAQ/Help	29	11.3%
index/table of contents/links	15	5.8%
welcome/homepage	15	5.8%
course description	8	3.1%
article	6	2.3%
course list	3	1.2%
diary, weblog or blog	3	1.2%
form	1	0.4%
forum/ interactive discussion archive	1	0.4%
job listing	1	0.4%
Total	257	100.0%

Table 5.8 (Appendix X) shows the votes for page 208. The page is a few paragraphs of someone's thoughts on the recent, unexpected death of his mother. There is a title at the top and an email link at the bottom for contacting the author, followed by a few links that suggest the site functions as an online support group. "Diary, weblog or blog", "personal

website", and "article" were the biggest vote-getters with 44.7%, 24.1% and 11.3%, respectively. "Forum/interactive discussion archive" (8.6%) may have gotten votes because of the prompt, "Questions?" and email link at the bottom of the page.

Table 5.8 Votes for page 208

Genre	# Votes	%
diary, weblog or blog	115	44.7%
personal website	62	24.1%
article	29	11.3%
forum/ interactive discussion archive	22	8.6%
speech	13	5.1%
none of the above	13	5.1%
poetry	2	0.8%
other instructional materials	1	0.4%
Total	257	100.0%

Table 5.9 Votes for page 219

Genre	# Votes	%
article	71	27.6%
none of the above	64	24.9%
speech	46	17.9%
welcome/homepage	42	16.3%
other instructional materials	15	5.8%
diary, weblog or blog	8	3.1%
FAQ/Help	4	1.6%
index/table of contents/links	2	0.8%
personal website	2	0.8%
forum/ interactive discussion archive	1	0.4%
product for sale/shopping	1	0.4%
search start	1	0.4%
Total	257	100.0%

Table 5.9 (Appendix Y) shows the votes for page 219. This page is one of several pages of a speech, but is formatted like an article or a homepage. This page had the lowest "high percentage" agreement of any page with "article" (27.6%). "None of the above"

followed with 24.9%. “Speech” and “welcome/homepage” received 17.9% and 16.3%, respectively.

There were four pages at what might be considered a borderline level of consensus (50 – 59%). A quick failure analysis shows how multiple aspects of a single page affect the level of single genre consensus. Page 1 (shown in Sample Screen Appendix Q) is clearly both an “index” (50%) and a “search start” (41%). Page 41, conveyed in a first-person style, contains an abstract, a table of contents, a bibliography and a link to the full paper that was delivered at a conference. Genres with the most votes were “article” (50%), “personal website” (22%), and “speech” (10%). Page 138 displayed one question and answer. “Forum/interactive discussion archive” (57%) and “FAQ/Help” (31%) dominated the responses. Finally, page 217 looks like an “article” (51%) but is a “speech” (30%).

In summary, the most commons causes for a lack of consensus on pages were: pages that didn’t seem to fit well in any genre in the palette, and pages with multiple aspects that seemed to fit into multiple genres. Occasionally, a page of one genre was formatted more like a page of another genre, which lowered consensus for the page.

Discussion

The aim of this study was to show that, given appropriate input from users, a palette could be constructed with genres, and those genres could be recognized with some level of accuracy by typical users. The Internet edu domain was used as an example webpage collection.

A major limitation of this study was the lack of control over participant selection and behavior. This was an anonymous study accessed over the Internet by participants. There was no way to monitor frivolous behavior, or to really know if the participants were of the target

group of faculty, staff and students in a university environment. There was no way to know how well, if at all, participants read the instructions, or the definitions of the genres which they were using to classify web pages. Participants could have participated more than once without detection. As in many studies, the sample was self-selected, and was probably not representative of the target population as a whole. Nothing is known about their institutional departments or academic disciplines. Given that genre is based on user community, this is a limitation of the study. For example, it is impossible to determine if participants with lower agreement with the majority may differ from the majority because they may belong to different demographic groups. Also, it cannot be determined if demographic differences could be a cause of a page's votes being almost evenly split mostly between two genres.

Both the corpus of web pages shown to the participants and the palette of genres are likely not completely representative of the edu domain. The restriction of one genre per page surely distorted the measurement of participants' perceptions of the pages' genres. The inability to modify genre decisions once they were made, may also have affected the results.

No measure of participants' confidence in their decisions was taken. It was impossible to distinguish whether disagreement by participants on a given page represented the existence of multiple genres from which participants chose, or whether they had trouble assigning any genre at all to a given page.

Also, it could be inferred from an inspection of the results that, at times, participants may have classified pages due to presence of specific words or phrases (e.g., home, welcome, frequently asked questions). In some cases, it seemed that these words may have been used to classify a page, even when the word or phrase appeared only in a link to another page (i.e., not the page being classified). Although keywords may be important in identifying genre,

reliance on them to the exclusion of other factors is not classification by genre. It is impossible to know what percentage of participant decisions were based on classification by keyword, rather than by genre.

Despite the limitations, overall participant agreement per page of 72% seems at the least to be acceptable, with an overall inter-participant agreement of 58% (and Cohen's kappa of .55, indicating "moderate agreement" (Landis & Koch, 1977)). Further, it must be remembered that genre represents typical, recurrent situations. It seems reasonable to assume that not all web pages are created out of typical recurrent, situations. Thus, users should not be expected to agree on the genre of all pages. However, 90% of the pages shown to participants in this study were classified into a single genre by a majority of the participants.

Several factors are hypothesized to have impeded participant agreement. As could be inferred from the failure analysis of the seven "non-consensus" pages, which were not classified into a single genre by the majority of the participants, some of the pages seemed to fit into multiple genres. For example, page 122 (Table 5.5) seemed to belong to both "article" and "product for sale/shopping". It presented excerpts of articles, and allowed you to buy the whole online book. As noted in the Results section, participant comments also mentioned the multiple genre per page phenomenon. This is consistent with previous research (e.g., Haas & Grams, 1998; Karlgren, et al., 1998). Soliciting more information than just a single genre per page could conceivably increase the level of participant agreement by allowing them to choose multiple genres.

In addition to some pages appearing "fuzzy" in terms of genre assignment, some of the genres seemed to be fuzzy, too. For example, one of the participants took issue with the definitions of "personal website" and "welcome/homepage" in his comments. In an email

exchange with the experimenter, he confirmed that his idea of a homepage was a “personal website”, and that a welcome page was the top-level entry point into a website, personal or otherwise. Further, a search of homepage definitions using Google uncovered both definitions: personal page and entry point. The classification results also support this dual definition: on the four pages in which “Personal Website” received a majority of the votes, “welcome/homepage” was second-highest once, third-highest twice, and fourth highest, once. (Disregarding “none of the above”, on these pages, “welcome/homepage” puts second to “personal website” twice, and third twice (behind “diary, weblog or blog” which also seems highly related to “personal website” according to the results). Having the terms “personal” and “homepage” in separate genre names seemed to cause disagreement among the users. Perhaps, combining the terms into “personal homepage” (suggested to be the “first truly digital genre” by (Dillon & Gushrowski, 2000)) should be worked into palette. Also, one participant commented, “Didn't know how to rate personal information in the context of institutional setting.” It is inferred that this comment was about a faculty member’s page that seemed to be written by the school, rather the individual. That page was rated by 63% as “personal website”, although 19% called it “none of the above”, possibly because of the quandary mentioned above. For all these reasons, “personal website” could be considered a fuzzy genre.

“Diary, weblog or blog” was ranked second on two of the four “personal website” pages, as identified by participants, and third on another. The relationship was somewhat reciprocal as “personal website” was ranked second on two of the “diary, weblog or blog” pages. “Diary, weblog or blog” had a similar reciprocal relationship with “forum/interactive discussion archive”. This is not surprising because blogs can serve both as personal journals

and as discussion sites (Herring, et al., 2004). The interrelationships between these three genres was alluded to by one of the participants, “it is sometimes harder to decide between blogs, forums and personal websites.”

Also, the “speech” genre seemed highly related to article. Eighty-five percent of the votes for the “speech” genre on the consensus pages and in which “speech” was not the consensus, were pages in which “article” was the consensus genre. There was not a reciprocal relationship in the data.

Several participants made specific comments about the “form” genre. Two referred to pages as forms in which “form” did not seem to capture the purpose of the document. “Survey” or “quiz” were suggested. One participant seemed to refer to “form” in terms of the HTML definition (e.g., “any site that had boxes to fill in or check was a form”). Thus, participants may have had disparate concepts of the “form” genre as well. A participant noted that “in some cases instructional material could also have been lumped under forms”, also suggesting the need for further palette refinement.

In terms of “false hits”, the “other instructional materials” genre was second only to “none of the above” (see Table 5.2). This suggests that “other instructional materials” is too broad, and may benefit from a name and/or description modification to narrow the genre’s scope. This is consistent with a participant’s suggestion “to make this category more clear or to split it out into more well-defined subcategories”. As noted earlier, it is not known whether participants read the definitions. In contrast to the last quote, another participant reported “I found the description of ‘Other Instructional Materials’ useful since I wasn’t sure what the category would have covered otherwise. I would have interpreted it much broader.” Perhaps some participants did not read the definitions.

The “article” genre was one of the least well recognized, both in terms of consensus (67.1%) and false hits (1.6%, see Table 5.2). The scope of “article” may need to be reduced by subdividing it into several more specific genres. “Article” suffers from the condition that its name alone is more general than any specific definition that might be given to it. As has been shown, one can write a specific definition for the article genre, and have people (more or less) agree on their classifications using it. However, the name “article”, by itself, most likely conjures up disparate notions in the minds of different people. In fact, comments by Swales (1990) indicate that he would not even consider it a genre.

> the English-speaking world…uses *names* to describe classes of communication that quite appropriately operate as higher-order categories than genres. One very common example is the *letter*. This useful term, of course, makes reference to the *means* of communication, but lacks as a class sufficient indication of purpose for genre status. The same observation holds for subsets of the class that refer to fields of activity such as business letters or official letters. It is only when purpose becomes ascribable that the issue of genre arises, as in begging letters, or letters of condolence. Category labels like *letters*…operate as convenient multigeneric generalizations. (p. 61)

Thus, “genres” proposed in this study such as “article” or “form”, despite their basis in user terminology, may need to be subdivided into more purposeful categories to be useful in the context of this research.

With all this evidence of the fuzziness of pages and genres, it may seem surprising that participant agreement was as high as 72%. Certainly, in future research, modifications could be made to the palette to attempt to reduce the fuzziness of some of the genres. Definitions to “personal website”, “diary, weblog or blog”, and “forum/discussion archive” could be changed to help distinguish them from one another. Perhaps “diary” and “blog” should be in separate genres. “Personal website” and “welcome/homepage” would benefit from being distinguished better. Given textual similarities between “speech” and “article”,

the elimination of the “speech” genre might be considered. More research may be needed to determine if the concepts of “fill-in form” and “HTML form” should be distinguished. The scope of “other instructional materials” and “article” might also be reduced.

Also, “course list” and “course description” could benefit from being better distinguished from each other, and perhaps “course description” is too broad, as evidenced by Table 5.2, and a participant’s comment, “a syllabus is different from a course description and a list of course descriptions is slightly different from a simple listing of available courses, but maybe I am splitting hairs.” Alternatively, as mentioned in the Results, a combination of the two genres might be effective. Certainly, some genres will need to be added: abstract, events schedule or list of events, survey, quiz, and calendar were among those suggested.

Measurement of participant perceptions could be improved but as always, there would be trade-offs between the quality of the results and effort needed to get them. Allowing participants the option of providing multiple genres seems necessary. Also, it would be helpful to have a confidence score for each rating. Allowing participants to change their answers might have some impact (especially according to this participant comment, “mislabled [sic] a few of the pages as blogs before I realized that there was a "forum/discussion" category”). Of course, a more controlled environment where participant behavior could be more closely monitored (e.g., time spent reading the genre definitions) would be beneficial, with the rating of a wider range of web pages.

It is interesting to compare this study’s genre palette with the two other user-derived genre palettes reported in literature (Karlgren, et al., 1998; Meyer zu Eissen & Stein, 2004). Despite the differences in scope (general Internet vs. the edu domain of this study), there

seems to be a close correspondence between much of the other two palettes and this one. See Table 5.10.

Table 5.10 Genres in the current palette compared to those of other user-derived palettes

The current study	Karlgren, et al, (1998)	Meyer zu Eissen & Stein (2004)
article	Journalistic materials Press: news, editorials, reviews, popular reporting, e-zines Reports Scientific, legal, and public materials, formal text Other running text	Article. Documents with longer passages of text, such as research articles, reviews, technical reports, or book chapters.
course description		
course list		
diary, weblog or blog		
FAQ/Help	FAQs	Help. All pages that provide assistance, e. g. Q&A or FAQ pages.
form		
forum/interactive discussion archive	Discussions Contributions to discussions, Usenet News material.	Discussion. All pages that provide forums, mailing lists or discussion boards.
index/table of contents/links	Link Collections	Link Collection. Documents which consist of link lists for the main part.
job listing		
other instructional materials		
personal website	Informal, Private Personal home pages.	Portrayal (priv). Private self-portrayals, i. e., typical private homepages with informal content
picture/photo		
poetry		
product for sale/shopping		Shop. All kinds of pages whose main purpose is product information or sale.
search start		
speech		
welcome/homepage	Public, Commercial Home pages for the general public	Portrayal (non-priv). Web appearances of companies, universities, and other public institutions. I. e., home or entry or portal pages, descriptions of organization and mission, annual reports, brochures, contact information, etc.
none of the above		

Table 5.10 Genres in the current palette compared to those of other user-derived palettes

The current study	Karlgren, et al, (1998)	Meyer zu Eissen & Stein (2004)
---		Download. Pages on which freeware, shareware, demo versions of programs etc. can be downloaded.
---	Interactive pages Pages with feedback: customer dialogue; searchable indexes.	
---	Other listings and tables	
---	Error Messages	

Note that the genres from the other studies were not validated by users in a reported study. However, the similarities (to the validated genres from this study) provide more evidence that some substantial amount of shared genre knowledge exists among users, even from different cultures (i.e., United States, Sweden, Germany).

In conclusion, the level of agreement about web genres among users looks promising, and there seem to be potential methods to explore for further increasing the level of agreement measured by experimental means.

The question at this point then becomes, which if any of these genres, would be useful in some way for improve information retrieval on the Web? For example, would the labeling of a webpage's genre in the search results presented to users allow them to make better or faster judgments regarding those search results? The next chapter reports on that study.

This page left intentionally blank.

CHAPTER 6

MEASUREMENT OF USER RELEVANCE JUDGMENTS OF GENRE ANNOTATED SEARCH RESULTS

The overall research question guiding this research concerns the value of genre as a document descriptor for improving web search effectiveness. This particular study addressed whether describing documents by genre in the list of search results improves web search effectiveness. However, rather than measuring outcomes of actual searches, the study focused on users' evaluation of search results in a simulated search environment. Users were presented with an information need, and then asked to evaluate a series of search results. The specific research questions were:

- Will people be able to make more better relevance judgments of search results using document descriptions that include genre?
- Will people be able to make faster relevance judgments of search results using document descriptions that include genre?
- Will users perceive document descriptions that include genre as an improvement over those that do not?

Throughout this discussion, the terms "surrogate", "summary", and "search result" may be used interchangeably with "document description" and "webpage description". Also, the term "full-text" may be used to indicate the actual webpage.

Each participant performed five tasks (one practice task and four actual tasks). For each task, the participant was presented with a written statement of the task's information need (modeled after Borlund's (2000) "simulated work task situation", which Borlund has suggested to be an adequate substitute for a participant's personal information need).

Working from that statement, the participant was asked to judge the usefulness of 20 web page descriptions (modeled after the format of Google search results). These descriptions were presented to the participant one at a time. Next, working from the same statement of information need as before, the participant was asked to judge, one at a time, the relevance of the 20 corresponding web pages. Thus, the study did not address how genre as a document descriptor will affect query formulation or reformulation, as participants did not develop their own queries, but rather, their judgments of search results.

Methods

Participants were recruited from the faculty and staff of a small, Southeastern private college. They were solicited through a couple of email messages to all faculty and staff (see Appendix Z). The sessions were conducted in unscheduled classrooms or seminar rooms on-campus at mutually agreed upon times.

Data Collected

Two primary measures were taken from each task: the time taken to judge each of the 20 document descriptions, and the 20 pairs of description and full-text usefulness judgments (one pair for each page). Participants' performance on tasks with genre-annotated descriptions was then compared with that of the tasks without genre annotations, on the basis of the time taken to judge the descriptions, and on the basis of the degree of correspondence between the ratings of the descriptions and the ratings of the corresponding pages. In order to more closely model a "real-life" web page evaluation, and improve the participant's ability to

judge the usefulness of the page, the pages which the document linked to were also available for viewing in considering the relevance of the document.

On the final task, participants were asked to verbalize the reasons for their actions and decisions during the evaluation of the first ten webpage descriptions and the first ten webpages of the task. These "think aloud" sessions were tape-recorded. (Timing of usefulness judgments made during the tasks incorporating the think aloud procedure was not used in the statistical comparison of the time to judge descriptions with and without genre-annotation.)

In addition to data collected during the tasks, participants answered demographic questions before performing the tasks, and afterwards, questions pertaining to their perceptions of the relevance judging process in relation to the genre annotation (or lack thereof) in the document descriptions. After each participant session, the links accessed during the evaluating of the webpages were obtained from the web browser's history file, and those pages were downloaded for future analysis.

Materials

Experimental Apparatus. A website was constructed so that the five tasks could be performed at a computer, under the supervision of the researcher. The opening screen contained instructions for the participant, and links to that participant's tasks (see Appendix AA). Clicking a link on the opening screen brought the participant to a screen with the description of that task's information need (see Appendix BB), and a "Begin" button that would bring up the search description/webpage rating screen and initiate the timing of the judgment for the first search description (see Appendix CC). Clicking a choice on the rating

screen would bring up the subsequent description or web page to be rated, or the “task complete” screen at the end of the task (see Appendix DD).

Participant Tasks. The statement of information need for each task was divided into “Background” and “Task” paragraphs similar to Borlund (2000), although Borlund used the labels “Simulated work task situation” and “indicative request”. It was felt that the labels used here would be perhaps more meaningful and less confusing to the experimental participants. See Figure 6.1 for the text of the information needs for the practice task and the four actual tasks.

Figure 6.1 Statements of information need for the tasks

Practice Task:

Background:
Your school is considering providing daycare for employees' children, but there are many details to be worked out. You have been asked to help prepare a recommendation as to how the school should handle the issue of daycare for employees' children.

Task:
Find information regarding whether it is a common practice for other schools to offer this, the cost to employees (if any), and any other policies involved in the establishment and maintenance of a daycare program.

Task 1:

Background:
You are a member of a taskforce charged with proposing actions to help prevent student attrition. You, in particular, have been assigned to investigate "early warning systems" at other schools. Early warning systems are typically designed to identify students who are at risk of dropping out or flunking out from college. They can identify students in a variety of ways (e.g., low grade, financial difficulties, professor identification, etc.) and include various interventions to fix the identified problems (e.g., counseling, career counseling, financial aid, etc.).

Task:
Find pages with specific information on comprehensive early warning system programs at other schools.

Task 2:

Background:

The General Education curriculum at your school includes discipline-specific introductory writing courses. Students must choose one such course, but it can be in any discipline in which an approved introductory writing course is offered. You are considering proposing changes to this component of the General Education curriculum. To aid you in your task, you decide to compare how other school's General Education programs expose students to introductory writing skills.

Task:

Find pages that contain course information for courses which cover introductory writing skills for freshmen.

Task 3:

Background:

You are an academic administrator who has developed a proposal for a "pre-law" minor at your school. Opponents of the proposal argue that many disciplines prepare students for law school, so that no one department should claim to be the school's one department for law school preparation. The proposal suggests that the political science department should administer the pre-law program. You want information about what departments are typically involved with these types of pre-law programs.

Task:

Find pages that indicate what departments at other schools administer the school's pre-law instruction.

Task 4:

Background:

You are interested in conducting scholarly research on grade inflation in higher education. Before you get started, you want to find about other research programs on the topic at other institutions, papers on the topic, or reliable statistics from other schools. You are seeking hard facts: how much inflation has occurred, and, if available, how many (and what types of) colleges and universities are affected. Although you will eventually want to look at opinions on whether grade inflation is really a problem, and hypotheses of the causes and solutions, right now you are interested in facts that describe the situation.

Task:

Find pages with such objective information, or that seem to link to such information.

An informal survey of people in faculty and administrative staff positions revealed that search needs for information in the edu domain often share a common theme: a comparison of what is being done at other schools with the "home" institution, and looking for new ideas by seeing what is done at other schools. This theme spanned diverse areas such as academic programs (at the overall program level, as well as the individual course level), research programs (overall programs, grant funding aspects, research areas, including identifying scholarly publications in an area), and administrative policies (e.g., benefits, tenure, sexual harassment, and grade inflation). Thus, faculty and administrative staff seem to have some similar, broad types of information needs concerning web search of other schools' pages.

The statements of information need were tailored to be understood by, and appropriate for, these types of participants. All were needs that both faculty and administrative staff might reasonably be motivated to perform. The statements were aimed at collecting information from other schools, without reference to specific schools. They covered a range of content-related areas (e.g., academic program requirements, descriptions of specific kinds of courses, school policies, aspects of research programs, etc.). Thus, task variance was minimized by restricting all tasks' information needs to those that might be reasonably expected to be pertinent for the user group, and by the use of tasks in which information is sought for other schools in general (as opposed to specific schools).

Webpage and Surrogate Collection. For each of the five tasks, it was necessary to collect 20 webpages and their corresponding descriptions. Once again, queries to Google, filtering out pages from anywhere but the edu domain, were used. Pdf and doc files were also excluded from the results. Each query delivered 100 results. Pages for each task were

collected with just one query, except for task 4, which required two. Keywords for the queries were chosen from each task's statement of information need. Figure 6.2 lists the keywords used for each query.

Figure 6.2 Queries for collecting web pages for the tasks

Practice Task - query 1:	daycare
Task 1 - query 1:	early warning student
Task 2 - query 1:	writing introductory OR freshman "general education"
Task 3 - query 1:	pre-law instruction
Task 4 - query 1:	grade inflation
query 2:	grade inflation statistics

Keywords were chosen on a trial and error basis in an attempt to balance several competing criteria for choosing pages from Google's search results. The goal for the percentage of relevant documents in each task document set was 50%, for two reasons. First, previous studies (e.g., Saracevic, 1969) have suggested that it is easier to judge non-relevant documents than relevant ones. If this is the case, the nature of the effects of genre annotation on the judgment process may differ, and an imbalance in the relative number of relevant and non-relevant judgments could make the comparison of the effects of genre on the two types of judgments less effective. Second, the government-sponsored Tipster program (Mani, et al., 1999), in which the majority of work in recent years involving document summary evaluation has been done, established a precedent by using a percentage in the range of 25% to 75% for all tasks. Of course, no actual percentage could be guaranteed *a priori*, as the study participants here were the ultimate judges of relevance.

Also, a range of the degree of relevance was desired: not only pages that were obviously relevant or not, but also pages that were somewhere in between. It seemed likely that the helpfulness of the genre-annotated descriptions might vary depending on the degree of relevance of the described page. At the same time, there was a concern about the overall duration of the experiment. “Grey area” pages would take longer to evaluate. Given that each participant would be making 200 decisions (five tasks, each with 20 descriptions and 20 webpages), there was concern about the ability of the participant to finish the session in a time frame that would not fatigue the participant. (After pilot tests ran long, sometimes over two hours, the experimental mechanism was modified to allow only 5 descriptions and 5 webpages to be shown in the practice task.)

Finally, it was also desired to have a range of genres corresponding to the palette validated in earlier work. It seemed likely that annotated descriptions might vary in usefulness depending on the genre of the page described, and that this possibility should be explored. At the same time, there was uncertainty involved in achieving this goal. A team of coders assigned genres to the web pages, and their actual decisions could not be predicted. (The coding procedure is described below.) Table 6.1 shows the number of pages of each genre that were assigned to each of the four tasks (excluding the practice task). The numbers add up to more than 80 (20 pages in 4 tasks) as some pages were assigned to multiple genres.

Table 6.1. Number of webpages per genre in each of the four experimental tasks

Genre Name	**1**	**2**	**3**	**4**	**Total**
article	2	3	2	8	15
course description		4	2		6
course list	1	14	4		19
diary, weblog or blog		1		2	3
FAQ/Help	9	2	4		15

Genre Name	1	2	3	4	Total
form	2				2
forum/interactive discussion archive	1			2	3
index/table of contents/links	1		3	6	10
other instructional materials	1	1		3	5
personal website	1	1	1	2	5
product for sale/shopping				1	1
welcome/homepage	7		10	1	18
	25	25	26	25	

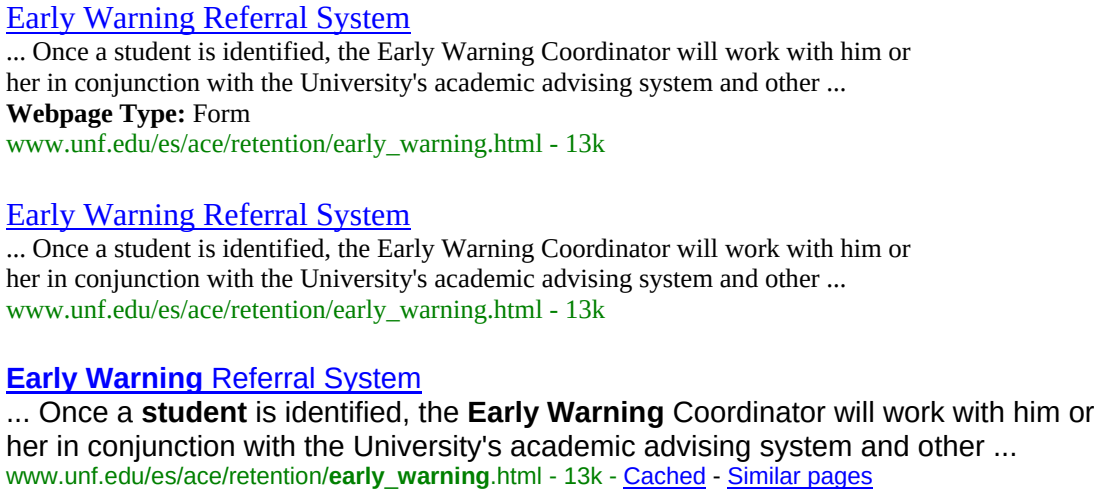
All pages and descriptions were stored in the researcher's website to provide a consistent presentation to the participants. Pages that the researcher could not make to display similarly to the original, were discarded from consideration. Often, this involved server-side functionality stored on the originating server (e.g., scripts for pre-loading menu images). The descriptions were presented as they appeared in Google's results pages, except with a slight difference in font, and no query keywords were bolded. Since participants did not submit queries in the tasks, it was felt that bolded keywords might be confusing, though it is acknowledged that bolded keywords are used under normal circumstances in search result evaluation. Appendix EE lists all the web page descriptions used in the study.

Genre Coding of Webpages. In order to assign genres for the labels that would be used in the genre-annotated web page descriptions, three independent coders were recruited by email: two graduate students from a school of information and library science, and a friend of the researcher with web search experience. Coders were emailed a URL to a webpage with links to the 80 pages to be coded. Links identified the pages by task and page number only. Coders were also emailed the list of 18 genre names and descriptions developed as reported in the previous chapter (see Appendix N). The coders were told to assign one or more genres to each page. They were allowed to code the pages in any order

they wished, and they could review and change their decisions as much as they wanted. When they were satisfied with their coding decisions, they emailed the page numbers and the assigned genre numbers back to the researcher.

A genre annotation was added to a page description only if it received at least two votes. Occasionally, the researcher voted to break a tie, or to increase the vote of a single genre that was one vote short of the genre(s) chosen by the majority of the three coders AND that was a genre, in the opinion of the researcher, that was at least as equally appropriate as the other genre(s) chosen. Descriptions with multiple genres listed them in alphabetical order, regardless of the number of votes. Figure 6.3 shows an example of the presentation of a page description. An un-annotated description did not have the line with the "Webpage Type" label. The name "Webpage Type" was chosen over "Genre Type" because it was thought that it would be more understandable to the participants.

Figure 6.3 Sample webpage description with genre-annotation, without genre-annotation, and the original Google description



Measurement Scale. The following prompt and scale were presented to participants to rate the webpages and descriptions:

Please rate the following web page/description according to how useful it is for resolving this search problem.

not useful at all ☐ **slightly useful** ☐ **somewhat useful** ☐ **highly useful** ☐

The decisions regarding the size and the labeling of the scale were not easy. The research literature seems to offer no clear choice for selecting the optimal method of measuring relevance in this study. The largest amount of conflicting arguments and evidence concerns the choice of the number of categories of the measurement scale: two (binary) or more. Studies grounded in what Schamber (1994) terms the "system view" of relevance have tended to favor binary relevance judgments, which facilitate the calculation of the traditional IR performance measures of recall and precision. However, an increasing number of researchers are acknowledging the inadequacy of binary judgments to measure the obviously varying degrees to which documents may be relevant to a specific need or query. Works suggesting this viewpoint are reviewed, and extended, in (Kekalainan & Jarvelin, 2002b). Other studies (e.g., Janes (1993)) have shown that judgments made on multi-category scales can tend to be bi-polar, anyway. Also, recent criticism of binary judgments (e.g., Sormunen (2002)) focused on "liberal criteria" for relevance. Such criteria, used in the TREC experiments (in which a perhaps trivial indication of topical relevance qualifies a document as relevant), are not essential components of binary judgments. Other criteria may be employed.

In the current study, it was the relevance to a user's information need that is being measured. That must be the case, as the focus of the study was to measure whether the addition of non-topical information to the search results can improve the relevance judgment

(which, then, must assess more than just topic). Thus, results from previous work assessing topical relevance may not be applicable here, because of the difference in the nature of the assessment.

Another significant difference between many past studies and this one is the studies' objects of assessment. They range from subject terms, titles, citations of varying content and format, and abstracts of varying length, to full-text. Few assess web search results and/or the full-text of web pages. Furthermore, sometimes the assessor judges both the surrogate and the full-text, but other times not. Sometimes the assessor is also the source of the information need, and other times the assessor is a "third party". These kinds of differences are well-documented (e.g., Mizzaro, 1997).

It must also be considered that the domain of the vast majority of these studies is that of scholarly research, which can be quite a bit more complex and ill-defined than the problems approached by many web searchers. Results garnered from such studies are of questionable generalizability to the current one.

Finally, the task presented to this study's participants was not really a search at all. It was essentially a set of 40 independent relevance assessments (with randomized ordering to obviate possible subject perception of search result/full-text pairings).

In summary, the literature (in general) upon which to base the choice of relevance scale is conflicting, and, of questionable applicability as well. (Appendix A shows past studies of document surrogate evaluation that are most similar to this proposed study.) Ultimately, a four-point scale was chosen to reflect the fact that relevance is not a simple binary decision. Having no mid-point, the scale forces the subject to make a choice, can

potentially detect shades of relevance, and can be collapsed to fewer levels, if that is deemed desirable.

Regarding the size and labeling of the scale, it was originally proposed to use the values of “highly relevant”, “somewhat relevant”, “somewhat irrelevant”, or “highly irrelevant” (as used in (Gordon & Pathak, 1999), a study that applied relevance assessment specifically to web pages). However, it was decided that “usefulness” would be an easier concept for participants to relate to than “relevance”. The labels eventually used (“not useful at all”, “slightly useful”, “somewhat useful” and “highly useful”) were pilot-tested, with participants reporting a good understanding and ability to use them. As in (Lan, 2002), no definition of usefulness was given to participants. The interpretation of usefulness was left to the participants.

Procedure

After greeting the participant, the participant was given the oral consent form (see Appendix FF), and the researcher answered any questions. The consenting participant was then given the entry questionnaire to complete (see Appendix GG). Upon completion of the entry questionnaire, the subject was seated at the computer with the opening screen displayed. The researcher explained what would happen during the session, and explained the instructions for performing the tasks. After answering any questions, the participant was asked to read the instructions on the screen, and when ready, to click the link to begin the practice task. It was emphasized to the participant that there was no wrong way to interpret the statement of information need. Whatever the participant decided was the object of this simulated search, it was the “right” thing. Participants were also told that their usefulness

ratings were not being compared to anyone else's, and that their rating was always the "correct" rating.

For the practice task, the document descriptions did not include annotations of document genre. Of the four experimental tasks, only two included genre-annotated descriptions. Each task contained either all genre-annotated descriptions, or all non-genre-annotated descriptions. The order of the tasks, and the order of which tasks included the genre annotations varied, for each participant.

After the practice task, subjects were asked about the clarity of the task scenario format (the information need), and the ease of rating descriptions and pages using the four-point scale provided. Any other questions were then answered. The participants were then told to work through the first three actual tasks, which could be accessed through links on the opening screen (Appendix AA). They were instructed to let the researcher know when they had finished those tasks, so that they could work together with the researcher on the final task. The think aloud process was referred to briefly at this time. Participants were told that they could ask questions or make comments at any time during the session.

After the participant completed the first 3 actual tasks, the researcher then explained the think aloud procedure. It was explained that sometimes the researcher might use prompts like "And what are you thinking now?" and "Why are you doing that?", following periods when the participant was silent. It was emphasized that the researcher was never questioning the participant's judgment, but simply trying to understand the reasoning behind the participant's decisions and actions. The participant was asked for permission to tape-record the think-aloud process, and was assured again that the identity of participants was not stored with the data. It was explained the think-aloud would be done only for the first ten webpage

descriptions, and the first ten webpages. For the tape recorder, the participant was asked to state the number of the description or page being rated (which was always displayed in the left hand side of the status bar at the bottom of the screen). The participant was also asked to verbally summarize what information the participant was seeking, when finished reading the statement of information need. If there were no questions, the participant was told to click the link for the final task when ready. At that time, the researcher started the recording, identified the participant by number aloud, and waited for the participant to state his/her interpretation of the information need. Then, the participant clicked the Begin button, and first ten descriptions were evaluated using the think aloud procedure. Then, the researcher shut off the tape recorder, and restarted it when the participant finished rating the last ten descriptions, and the first webpage to be evaluated was displayed. The participant and researcher followed the think aloud procedure until the first ten pages were rated. Then the researcher shut off the tape recorder, and the participant finished rating the remaining ten webpages.

Upon completion of the final task, the researcher conducted an exit interview with the participant. (See Figure 6.4 for a script of the exit interview.) After the interview, the participant was asked to sign a form acknowledging receipt of compensation, given $10, and thanked. When the participant left, the researcher copied the history file of the web browser for later downloading of pages whose links were clicked on by the participant during the experiment.

Figure 6.4 Exit interview script

The following questions pertain only to the web pages summaries (and not the actual web pages) that you rated for usefulness.

1. How easy or hard did you find it to rate the usefulness of the summaries and why?

2. What parts of the web summaries did you use to determine the summaries' usefulness?
3. Did you notice the "Webpage Type:" label in some of summaries?
   (If no, depending on time, ask: Do you ever try to figure out the type of web page just from looking at the summary? and follow-up appropriately; or just skip to question 8.)
   Was the label understandable to you? What did you think it meant?
4. Was the label helpful? Why or why not? Can you think of specific examples in which the label was helpful? Not helpful?
5. Did you use the label to help you make any usefulness decisions?
6. Was the label more helpful for indicating usefulness or lack of usefulness? Why?

These questions pertain to the actual web pages that you rated for usefulness.

7. When you were rating the web pages, how often were you aware that you had seen a summary describing the web page?
8. How much did you remember about a document's summary?
9. Did you remember your rating, or the type of the document?

Regarding any aspect of the entire study, do you have other comments?

Results

Thirty-two people, 18 faculty and 14 administrative staff completed the experiment during the time period of November 2, 2004 through December 1, 2004. Logistics made the precise recording of session times difficult, but the researcher estimates that the average time taken to complete the session was approximately 1 hour, 45 minutes. (For example, participants were mostly scheduled in two-hour back-to-back time slots. If a session was running into the next time slot, sometimes the next participant would read the oral consent form and complete the entry questionnaire while the first participant was finishing his/her session.)

Participants

Demographic information collected from participants is shown in Table 6.2.

Table 6.2 Participants' demographic information

	Faculty	Staff	Average	Min	Max
Average Years as Faculty/Staff	16.5	10.9	14.3	1	43
Average Hours/Week using the web	11.9	25.9	18	1	80
Average Hours/Week searching the web	4.8	11.3	7.7	0	70
Average Job-related Hours/Week web searching	3.7	7.6	5.4	0	50

(Faculty's hours as Staff were minimal, and excluded from the Faculty hours calculation.)

The entry questionnaire also asked participants if they had ever searched for information on topics that, unbeknownst to them at the time, were the topics of the tasks they would be performing. Of the 32 participants, 15 reported previous searching for one or more of the topics. Table 6.3 reports the number of participants by topic.

Table 6.3 Number of participants that reported experience with task topics

Task	Topic	# Reporting Experience
Practice	provision of daycare by colleges and universities	6
1	"early warning systems" for preventing student attrition	3
2	college composition courses	7
3	pre-law curricula	1
4	grade inflation	8

The data confirm the idea that people can have many different reasons for searching the same topic. The reasons that participants gave for these previous searches were diverse. For the daycare topic, three of the seven had children, one was a reference librarian, one has a job assisting adult students, one worked at a university that had a web-based clearinghouse of local daycares, and one English professor said that it was a topic that students were interested in writing about. Only one mentioned comparing daycare situations at other

academic institutions. Of the three with search experience in the early warning system topic, two participants actually work with early warning system policy, and another was a reference librarian. For the college composition course topic, one searches specific schools in order to approve transfer credits, another was searching at specific local institutions for general education requirements, one was taking a college composition course and wanted to compare it to that of other schools, two faculty were interested in new ideas for degrees or programs, one taught legal research and writing and compiled teaching resources for adjunct faculty, and one was a reference librarian. Only one person had experience searching pre-law curricula: she runs the continuing education paralegal program at her institution. For the grade inflation topic, of the eight with that experience, three were members of academic committees studying the topic, two were professors searching out of their own personal interest (one used the information to influence her own grading scale), one was a statistics professor looking for a dataset to use in class, one was looking for information to give to senior management, and, of course, one was a reference librarian.

Exit Interview

The exit interview was conducted with each participant after the final task was completed. The main focus of the interview was to get their feedback on the usefulness of the "Webpage Type:" label used in half of the participants' tasks. See Figure 6.4 for the exit interview script. Following is a summary of participants' responses to the questions.

Note that the sessions often ran close to, or over the scheduled 2-hour session time. Thus, there was not always time to ask for clarification of answers, or to ask follow-up questions which some answers may have prompted.

The purpose of the first two questions was really to encourage the participants to volunteer perceptions of the genre annotation without specifically prompting them.

Question 1: How easy or hard did you find it to rate the usefulness of the summaries and why?

Participants' perceptions of the difficulty of judging the summaries ranged widely from "fairly difficult" to "sometimes hard, sometimes easy" to "pretty easy". Regardless, many felt that the level of difficulty was typical of web searching. There were a few exceptions. One mentioned that having keywords highlighted (like normal Google summaries) was more informative and descriptive. Two thought that the summaries were shorter than is typical. (Given that the summaries were taken verbatim from Google, it is possible that these two primarily use other search engines.) One said that the task was harder because the results were not presented in a list, as search results typically are. Three people said that the assigned tasks made searching harder than usual, either because they lacked "background" or "insider knowledge" on the topic. For example, one described his typical searches as having "less nebulous" concepts (e.g. specific science terminology). Thus, not all of the tasks may have been tailored well enough for all of the participants (a recommendation by Borlund (2000) for developing an effective "simulated work task situation"). One also noted that it was easier to look for "finite things", examples given being "course descriptions" and "pre-law or not". Finally, a couple participants also noted that titles and the presence of keywords in summaries could be misleading at times.

Several recurring themes around the summaries were evident. Most felt that not enough information was provided. Many expressed difficulty with interpreting the partial sentences taken out of context. They noted that the judgments required a lot of guessing and interpretation. This is consistent with the fact (reported in more detail later in the Results

section) that about 65% of the judgments changed from summary to webpage. Three participants volunteered the “Webpage type:” label as being helpful for rating the summaries in response to this question.

Question 2: What parts of the web summaries did you use to determine the summaries’ usefulness?

The title, the description, the URL (or “web address”) and keywords were mentioned almost universally. However, there seemed to be a wide variety in the sequence and the amount of importance placed on each of those. For example, one person reported looking at the URL, only if he was still confused after looking at the description and the title. Another said he would “go crazy” if the URL were not in the summary. The URL was mentioned as providing information about the name of the school, the source of the information, the web domain of the page, and the type of page (“personal” being given as an example of page type). Finally, a couple of people mentioned the strategy of looking for things to help them eliminate the page from further consideration. Aspects of a summary that denoted a lack of relevance (e.g., the presence of words not related to the topic) were seen as helpful. Two more people (other than those mentioning the “Webpage type:” label in response to question 1) volunteered that the label was helpful in determining the usefulness of the summaries.

Question 3: Did you notice the “Webpage Type:” label in some of the summaries?

Participants were intentionally not told that the label would be in the summaries of two of there tasks. The rationale for this was that it was desired that the judging process be as close to the typical experience as possible. It was thought that drawing attention to the label might cause the participants to give it more consideration than they normally would. As it turned out, only 17 of the 32 participants reported noticing the label, in response to this question. However, two of those “did not use it at all” or “didn’t pay any attention to it”.

Analysis of the data comparing the performance of those noticing the label with those who didn't shows little difference between the two groups. Twelve participants who reported noticing the label were able to recall specific genre labels. The most frequently recalled labels were article (5), diary/blog (4), personal (4), and welcome/homepage (4). Recalled to a lesser extent were FAQ (3), index/table of contents/links (3), discussion forum (2), course listing (2) and form (1). Note that not everyone used exactly the same terminology in recalling the labels. Labels not mentioned by anyone were course description (which may have been conflated with course listing), instructional materials and shopping (of which there was only one page).

Question 4: Was the label understandable to you? What did you think it meant?

Everyone (17) who noticed the label, and who did not indicate ignoring it (the two just described above), reported that they felt like they understood it. One noted that the more specific labels were more understandable and useful, e.g. personal website vs. the "too general" welcome/homepage.

Information pertaining to questions 5 through 7 was frequently combined by participants, and pertains to more than one of these questions. For ease and economy of exposition, responses pertinent to these questions will be discussed together.

Question 5: Was the label helpful? Why or why not? Can you think of specific examples in which the label was helpful? Not helpful?

Question 6: Did you use the label to help you make any usefulness decisions?

Question 7: Was the label more helpful for indicating usefulness or lack of usefulness? Why?

Twelve people specifically expressed that the labels were helpful. One reported that the labels "gave you an idea of what the document was and the format." Similarly, another

two reported, “yes, they helped me figure out what I was looking at” and “they were helpful in sorting things out.” One was enthusiastic about the labels: “Oh, this is really good.” More specifically, another said the labels “helped distinguish usefulness, credibility” (citing blog as an example of a lack of credibility). One said that the label should be more prominent.

Overwhelmingly, those who had an opinion on the issue mostly felt the labels were better for indicating lack of usefulness, than for indicating usefulness. Only two thought that the labels were equally helpful for indicating either. No one reported thinking that the labels were better for indicating usefulness. One participant explained it this way, “the right type doesn’t guarantee usefulness, but the wrong type guarantees a lack of usefulness.”

Frequently, participants reported judging a summary as not useful based on the label, and then later being surprised by the actual webpage’s usefulness. This happened most frequently with the labels of personal website and diary/blog. Many participants seemed to regard these labels as inherently not useful. Additional labels mentioned in this regard were discussion archive and FAQ. Some participants seemed more discerning: “I wouldn’t look at a FAQ unless I was looking for an answer”. Another noted that decisions were not based solely on the labels.

There was time to explore general use of page types with six of the participants who reported not noticing the presence of the labels in response to question #3. Two mentioned that they didn’t actively look for page types, but may have used them if they noticed them. One reported not thinking about page types, and couldn’t remember any or using any. Types mentioned by the other three participants included magazine, blog, text, personal web page, article, abstract and “report for the task force” (referring to one of the experimental tasks’ pages). Their answers differed from the overall group in estimation of relative ability for

indicating usefulness/un-usefulness. One said that page types were equally good for both. Another said that the types were better for indicating usefulness. A third said they were good for neither, because type information "overlaps" with the URL.

Question 8: When you were rating the web pages, how often were you aware that you had seen a summary describing the web page?

The majority of the participants realized the connection between the summaries and the pages. A few assumed the connection for all the pages or most of the pages, even if they didn't specifically recognize it as often as that. Seven reported noticing the connection over time, two of them not until the final task which involved the think aloud procedure. One participant reported not noticing any connection.

Question 9: How much did you remember about a document's summary?

Specific aspects remembered from the summaries were mostly names (of schools, people, departments, publications, etc.) and titles. Also mentioned were idiosyncratic phrases. One person reported remembering the URL. In terms of how much was remembered, one said, on a scale of 1 to 5, she rated it a 1 (i.e., she remembered very little).

Question 10: Did you remember your rating, or the type of the document?

Half the participants reported remembering little or nothing about their summary ratings. Three reported having a "general sense" of their ratings. Eleven reported experiencing surprise or personal validation when viewing the pages. Stated one, "On a few, I remembered enough to have be happy or disappointed when I saw the page." A couple reported a tendency to remember their extreme ratings. One participant concluded that she learned to "give more web sites a chance; the summary is less accurate than I previously thought."

No one specifically reported remembering the "Webpage Type:" label from a summary when viewing the corresponding page, though participants seemed to have a general sense of page types, e.g., "that Chronicle article turned out to be a book review."

Relevance Decisions: Correspondence of Summary to Full-text

Excluding the practice tasks, the thirty-two participants made a total of 2,560 pairs of relevance decisions. Due to a small number of data collection problems, six pairs were unusable, so the total number of pairs in the data analysis is 2,554. Table 6.4 summarizes the decisions made by all the participants.

Table 6.4 Usefulness – all participants

	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	363	107	53	85	608	59.7%	0.77	0.77
2	282	178	115	125	700	25.4%	0.92	0.12
3	218	172	128	185	703	18.2%	1.13	(0.60)
4	126	105	82	230	543	42.4%	1.23	(1.23)
Total	989	562	378	625	2554	35.2%	1.01	(0.21)

In Table 6.4, the rows are the ratings of the summaries while the columns on the left side of the table are the ratings of the actual webpages. The Total column shows the distribution of ratings of the summaries. For example, of the 2,554 ratings of summaries, 608 were 1s (not useful), 700 were 2s (slightly useful), 703 were 3s (somewhat useful) and 543 were 4s (highly useful). The Total row gives the distribution of ratings of the actual webpages. 989 were 1s, 582 were 2s, 378 were 3s and 625 were 4s. It is interesting to note that the distribution of final judgments became more bi-polar (relatively more 1s and 4s), as described in much of the research on relevance judgments (Greisdorf & Spink, 2001).

The 4 by 4 matrix of 16 numbers shows how the ratings of the summaries relate to the ratings of their corresponding webpages. For example, of the 608 summary ratings that were 1s, 363 of those summaries' webpages were also rated as 1s. Thus, for the 363 summaries rated as not useful, the rating did not change when the corresponding webpage was rated. The percent of summary ratings that did not change is the "indicativity", which is 363/608, or 59.7%, as shown in the Indicativity column. Similarly, 178 of the 700 summaries that were rated a 2, described a web page that was also rated a 2, for an indicativity of 25.4%. Indicativity (McLellan, et al., 2001; Marcus, et al., 1978) is a measure of how well the rating of the summary predicts the actual relevance of the web page. McLellan, et al. (2001) described the use of indicativity as a measure:

> It was essential to establish a baseline for these measurements, the ground truth. In this case, the ground truth was taken to be a subject's relevance assessment based on the complete text of a document. The rationale for this choice is that once a user has seen the full text there is no more information to be presented, and so a more accurate judgment cannot be made. Therefore, if a judgment changes after having seen the full text, it is clear that the summary did not contain enough information to make a correct judgment. (p. 107)

Overall indicativity for the 2,554 rating pairs is 35.2%. That means, given a summary rating, there was a 64.8% chance that the rating of the web page would differ. Note that an indicativity of 100% would mean that the ratings of the summaries perfectly align with the ratings of the webpages. Also note that, on a four point scale such as the one used here, an indicativity of 25% is the level of pure chance (i.e., rating the summaries and web pages randomly).

Indicativity by itself is not informative of the magnitude of the change between the ratings of the summary and page. It simply indicates the percentage of ratings that did not change. The column to the right of Indicativity in Table 6.4 is the "Absolute Change". It is

the average of the absolute amount of change between all the rating pairs. Table 6.4 shows that the average change over all 2,554 rating pairs is 1.01. Note that Absolute Change ignores the sign of the change. Thus, a change from 1 to 2 is the same as a change from 4 to 3, The absolute value of difference in both cases is one. Thus, Absolute Change gives a little more information than Indicativity, although it should be noted that ratings at the extremes of a scale have potential for greater changes then the middle points of a scale (because extreme points are farthest away from the opposite extreme). McLellan, et al. (2001) report this Absolute Change as "variance".

A weakness of Absolute Change is that it does not take the direction of the change in to account. It conveys only the size of the change. Thus, another useful measure is the column to the right of Absolute Change: the Relative Change. Relative Change is the simple average of the differences between rating pairs. For example, a change from 1 to 2 and a change from 4 to 3 averages to a Relative Change of 0, since a change of positive one and negative one average to zero. Table 6.4 shows that the Relative Change for the rating pairs is -.21, meaning when all the changes are averaged, the ratings of the web pages average .21 less than the ratings of summaries. Note that Relative Change and Absolute Change are always equal for the endpoints (1 and 4) of the scale. The measures of Absolute Change and Relative Change can help to give us a better picture of the nature of the rating changes in a set of data. However, they are solely for descriptive purposes, rather than for use in tests of statistical significance.

Although multi-level relevance scales are thought to be more appropriate for measuring the multidimensionality of user relevance judgments (e.g., Spink, Griesdorf & Bateman, 1999), it was desired to also provide results to allow comparisons with studies

using binary relevance scales. Given that different user groups may have different "breakpoints" between relevance and non-relevance (Janes, 1991b), the best way to collapse the ratings into a binary scale was not obvious.

Table 6.5 Four ways to collapse the relevance decisions into a binary scale

	Full-text Rating					
Summary Rating	1	2,3,4	Total	Indicativity	Absolute Chg	Relative Chg
1	363	245	608	59.7%	0.40	0.40
234	626	1320	1946	67.8%	0.32	(0.32)
Total	989	1565	2554	65.9%	0.34	(0.15)

	Full-text Rating					
Summary Rating	1	4	Total	Indicativity	Absolute Chg	Relative Chg
1	363	85	448	81.0%	0.19	0.19
4	126	230	356	64.6%	0.35	(0.35)
Total	489	315	804	73.8%	0.26	(0.05)

	Full-text Rating					
Summary Rating	1,2,3	4	Total	Indicativity	Absolute Chg	Relative Chg
123	1616	395	2011	80.4%	0.20	0.20
4	313	230	543	42.4%	0.58	(0.58)
Total	1929	625	2554	72.3%	0.28	0.03

	Full-text Rating					
Summary Rating	1,2	3,4	Total	Indicativity	Absolute Chg	Relative Chg
12	930	378	1308	71.1%	0.29	0.29
34	621	625	1246	50.2%	0.50	(0.50)
Total	1551	1003	2554	60.9%	0.39	(0.10)

Table 6.5 shows four ways to collapse the data from Table 6.4 into ratings of relevant and not relevant. The first three methods of collapsing the surrogates and full-text judgments were used by Saracevic (1969). The first set of numbers seems the most intuitive way of the methods: partially relevant ratings (2 and 3) are combined with the relevant ratings (4). The second set of numbers ignores the partial relevance ratings. The third set combines the partial relevance ratings with the non-relevant ratings (what might be considered a method enforcing "strict" relevance criteria, according to Gordon and Pathak (1999)). The fourth method splits

the 4-point scale down the middle (a “lenient” encoding of relevance (Gordon & Pathak, 1999)).

The collapsed tables in Table 6.5 seem to offer no surprises. Predictably, the indicativity of the most relevant rating increases when multiple categories are combined with it. Also, predictably, overall indicativity increases when the partial judgments are ignored. This result parallels the finding by Saracevic (1969), highlighting the relative instability of the partial judgments. The lowest level of relevance increases in indicativity when multiple categories are combined with it. Mathematically, one can generally expect indicativity to increase as the number of rating levels is decreased. Also note that Absolute and Relative Change will decrease as the maximum change is reduced from three to one. As no real advantage to collapsing the data has been observed, the remainder of the results will be presented uncollapsed.

Table 6.6 shows that there was little appreciable difference between the performance of faculty and staff.

Table 6.6 Usefulness – faculty vs. staff

FACULTY	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	209	60	30	40	339	61.7%	0.71	0.71
2	169	111	78	64	422	26.3%	0.89	0.09
3	122	96	69	106	393	17.6%	1.13	(0.60)
4	56	63	42	124	285	43.5%	1.18	(1.18)
Total	556	330	219	334	1439	35.6%	0.97	(0.20)

STAFF	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	154	47	23	45	269	57.2%	0.85	0.85
2	113	67	37	61	278	24.1%	0.98	0.17
3	96	76	59	79	310	19.0%	1.12	(0.61)
4	70	42	40	106	258	41.1%	1.29	(1.29)
Total	433	232	159	291	1115	34.6%	1.06	(0.22)

Table 6.7 compares relative indicativity of the genre-annotated description and the non-genre annotated descriptions. Genre-annotated descriptions seem to allow users to better identify non-relevant web pages (63.5% vs. 55.8% indicativity for the 1s), but with an overall loss of indicativity on the other three rating levels, such that total indicativity is virtually unchanged (35.3% vs. 35.1%). For each participant, the total indicativity for trials with genre annotation and for trials without genre annotation were computed. A paired t-test showed no significant difference between the means of total indicativity with and without genre annotation, $t(31) = -.011$, $p = .91$. A similar test to determine the significance of the difference in the indicativity for the non-relevant web page descriptions would not be statistically meaningful as the number of descriptions rated non-relevant varies by participant (whereas, for the previous test, the number of descriptions that participants rated with and without genre annotation did not vary, except very slightly, due to the six missing datapoints explained earlier).

Table 6.7 Usefulness – by presence of genre annotation

W/ GENRE	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	195	53	26	33	307	63.5%	0.66	0.66
2	147	83	59	66	355	23.4%	0.95	0.12
3	102	89	61	92	344	17.7%	1.12	(0.58)
4	70	49	41	113	273	41.4%	1.28	(1.28)
Total	514	274	187	304	1279	35.3%	1.00	(0.24)

PLAIN	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	168	54	27	52	301	55.8%	0.88	0.88
2	135	95	56	59	345	27.5%	0.90	0.11
3	116	83	67	93	359	18.7%	1.14	(0.62)
4	56	56	41	117	270	43.3%	1.19	(1.19)
Total	475	288	191	321	1275	35.1%	1.02	(0.19)

Table 6.8 compares the relative indicativity of the two summary types by task, not including the practice task. The overall indicativity of task 2 is slightly less than the others, consistent with the investigator's impression that task 2 was the more difficult of the tasks and the assumption that time taken to complete a task correlates to task difficulty. The summaries of tasks 1 and 4 seem to be better at indicating non-relevance than those of tasks 2 and 3. However, the summaries of task 3 seem to be better at indicating relevance than those of the other tasks.

Table 6.8 Usefulness – by task

TASK 1	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	94	26	12	10	142	66.2%	0.56	0.56
2	64	50	27	27	168	29.8%	0.86	0.10
3	51	45	43	48	187	23.0%	1.04	(0.53)
4	30	37	27	47	141	33.3%	1.35	(1.35)
Total	239	158	109	132	638	36.7%	0.96	(0.30)

TASK 2	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	79	28	14	19	140	56.4%	0.81	0.81
2	86	60	26	29	201	29.9%	0.85	(0.01)
3	71	48	24	34	177	13.6%	1.27	(0.88)
4	32	24	16	46	118	39.0%	1.36	(1.36)
Total	268	160	80	128	636	32.9%	1.05	(0.32)

TASK 3	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	89	33	15	38	175	50.9%	1.01	1.01
2	59	33	24	46	162	20.4%	1.08	0.35
3	44	35	31	54	164	18.9%	1.08	(0.42)
4	21	13	27	78	139	56.1%	0.83	(0.83)
Total	213	114	97	216	640	36.1%	1.01	0.08

TASK 4	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	101	20	12	18	151	66.9%	0.65	0.65
2	73	35	38	23	169	20.7%	0.93	0.07
3	52	44	30	49	175	17.1%	1.13	(0.57)
4	43	31	12	59	145	40.7%	1.40	(1.40)
Total	269	130	92	149	640	35.2%	1.02	(0.30)

Table 6.9 shows the relative indicativity of the summary by types by task and by presence of genre annotation. For task 1, the pattern of results is similar to that of all four tasks combined (Table 6.7). Non-relevance has slightly higher indicativity (by 6.6%) with the genre annotation, but the other three rating levels are lessened, so much so in this case that, overall, the absence of a genre annotation made those summaries more indicative than those with genre annotation by 4.4%. For task 2, indicativity for genre-annotated summaries was slightly higher for non-relevance (by 4.6%), and partial relevant judgments (by 4.6% and 4.5% respectively, for 2s and 3s), but slightly lower for relevant judgments (by 2.1%). For overall indicativity of genre-annotated summaries, task 2 had the best relative performance, with 3.7% higher indicativity than the non-annotated ones.

Results for task 3 follow a different pattern. Genre-annotated summaries were more indicative of relevance and non-relevance, by 5.6% and 15.7%, respectively. However, the partial judgments of the annotated summaries are less indicative, to the extent that, overall, they are essentially equally indicative as the plain summaries (.4% lower). For task 4, genre-annotated summaries are about equal in indicativity to the plain ones for relevance and non-relevance, while genre annotation hurts the indicativity of the "slightly useful" (2) ratings, while helping the indicativity of the "somewhat useful" (3) ratings.

Statistical tests for the effect between each task and the presence of genre annotation were done as follows. For each participant, the total indicativity for each task was computed. The two-sample t-test was used rather than the paired t-test because for any given task, each participant either rated summaries all with the annotation, or all without. For each task, a t-test showed no significant difference between the means of total indicativity with and

without genre annotation: for task 1, $t(29.5) = 1.11$, $p = .28$; for task 2, $t(30) = –0.81$, $p = .43$; for task 2, $t(29.5) = 0.08$, $p = .94$; for task 3; and for task 4, $t(29.8) = –0.61$, $p = .55$.

Table 6.9 Usefulness – by task and presence of genre annotation

1 W/ GENRE

TASK 1 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	52	13	6	4	75	69.3%	0.49	0.49
2	34	19	11	11	75	25.3%	0.89	(0.01)
3	27	27	20	22	96	20.8%	1.07	(0.61)
4	15	22	17	19	73	26.0%	1.45	(1.45)
Total	128	81	54	56	319	34.5%	0.98	(0.40)

PLAIN

TASK 1 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	42	13	6	6	67	62.7%	0.64	0.64
2	30	31	16	16	93	33.3%	0.84	0.19
3	24	18	23	26	91	25.3%	1.01	(0.44)
4	15	15	10	28	68	41.2%	1.25	(1.25)
Total	111	77	55	76	319	38.9%	0.93	(0.20)

2 W/ GENRE

TASK 2 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	40	19	5	4	68	58.8%	0.60	0.60
2	49	36	12	16	113	31.9%	0.82	(0.04)
3	29	22	13	17	81	16.0%	1.20	(0.78)
4	19	10	7	22	58	37.9%	1.45	(1.45)
Total	137	87	37	59	320	34.7%	0.98	(0.35)

PLAIN

TASK 2 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	39	9	9	15	72	54.2%	1.00	1.00
2	37	24	14	13	88	27.3%	0.88	0.03
3	42	26	11	17	96	11.5%	1.32	(0.97)
4	13	14	9	24	60	40.0%	1.27	(1.27)
Total	131	73	43	69	316	31.0%	1.11	(0.30)

3 W/ GENRE

TASK 3 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	48	10	7	16	81	59.3%	0.89	0.89
2	28	14	14	28	84	16.7%	1.17	0.50
3	26	19	9	26	80	11.3%	1.21	(0.56)
4	15	4	12	44	75	58.7%	0.87	(0.87)
Total	117	47	42	114	320	35.9%	1.04	0.01

PLAIN

TASK 3 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	41	23	8	22	94	43.6%	1.12	1.12
2	31	19	10	18	78	24.4%	0.99	0.19
3	18	16	22	28	84	26.2%	0.95	(0.29)
4	6	9	15	34	64	53.1%	0.80	(0.80)
Total	96	67	55	102	320	36.3%	0.98	0.14

4 W/ GENRE

TASK 4 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	55	11	8	9	83	66.3%	0.65	0.65
2	36	14	22	11	83	16.9%	0.96	0.10
3	20	21	19	27	87	21.8%	1.01	(0.39)
4	21	13	5	28	67	41.8%	1.40	(1.40)
Total	132	59	54	75	320	36.3%	0.99	(0.21)

PLAIN

TASK 4 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	46	9	4	9	68	67.6%	0.65	0.65
2	37	21	16	12	86	24.4%	0.90	0.03
3	32	23	11	22	88	12.5%	1.24	(0.74)
4	22	18	7	31	78	39.7%	1.40	(1.40)
Total	137	71	38	74	320	34.1%	1.06	(0.40)

Tables showing decisions by individual genres by presence of genre can be found in Appendix HH. The fact that the number of pages per genre was not evenly distributed across tasks prevents meaningful tests of significance for the effects of genre annotation on specific genres.

Effect of Genre Annotation on Summary Decision Times

Table 6.10 shows the average number of seconds it took to judge the summaries by task and by task order (which was randomized). Times for the fourth task were not considered as the think aloud procedure affected those measurements. Summary decisions in tasks 2 and 4 took longer than tasks 1 and 3, 13.3 and 13.8, vs. 11.6 and 11.2. However, despite the short practice task, there seemed to be a task order effect in which the first task took longer than the last two, 14.3 vs. 11.9 and 11.3. Averaging decision times without the first task shows that task 2 clearly took longer than the other four tasks, which is consistent with the investigator's observation that task 2 seemed to be the most difficult of the tasks.

Regarding the stability of the average decision times per task, each number is the average of 100 to 220 decisions (as the randomization process did not allocate the tasks across the task orders evenly).

Table 6.10 Average summary decision times by task & task order (in seconds) (excluding the think-aloud task)

	Task				
Task Order	1	2	3	4	Average
1	14.2	12.3	11.9	17.5	14.3
2	11.0	16.2	11.2	10.0	11.9
3	9.5	12.4	10.6	11.9	11.3
Average	11.6	13.3	11.2	13.8	12.5
Average (2nd & 3rd only)	10.5	14.0	11.0	11.3	11.6

Overall, it seems that decisions on genre-annotated summaries took longer, 13.0 vs. 12.0 seconds (not shown in any table). Without taking any potential task effect into account, a paired t-test showed no significant difference betweens the participants' average summary decision times by the presence of the genre annotation, $t(31) = -1.67$, $p = .11$.

To investigate a possible task effect, statistical tests for the effect between each task and the time to rate the summaries were done as follows. For each participant, the average time to rate a summary for each task was computed. The two-sample t-test was used rather than the paired t-test because for any given task, each participant either rated summaries all with the annotation, or all without. For tasks 1 through 3, a t-test showed no significant difference between the means of average summary decision time with and without genre annotation: for task 1, $t(15.6) = .03$, $p = .97$; for task 2, $t(19) = –0.21$, $p = .83$; for task 3, $t(21) = 0.21$, $p = .83$. For task 4, the difference was statistically significant, but the direction indicated that genre annotated summaries took longer to judge for task 4, $t(20.5) = –2.1$, $p = .0482$. Table 6.11 shows the average summary decision time by task and by the presence of genre-annotation in the summary.

Table 6.11 Average summary decision times by task & genre annotation (in seconds) (excluding the think-aloud task)

Genre Annotation	Task 1	2	3	4	Average
With	11.6	13.5	11	15.5	13.0
Without	11.7	13.1	11.3	11.9	12.0

Table 6.12 shows the average summary decision times by task, presence of genre annotation and rating level. Times which are lower for the genre-annotated summaries are in bold. Genre-annotation seems to have helped a great deal in identifying “not useful” (1) pages in task 2, as decision times were quicker, 11.5 to 14.9 seconds. Genre-annotation also seemed to aid slightly faster decisions on both “not useful” (1) and “highly useful” (4) pages in task 3, 10.4 to 11.1, and 10.2 to 11.0, respectively. Surprisingly, genre-annotation helped the most with “slightly useful” (2) decisions in task 1, 11.9 to 15.7 seconds. Genre-annotation

allowed marginally faster "somewhat useful" (3) decisions for task 2, 12 to 12.7 seconds. However, most of the comparisons in Table 6.12 indicate that participant decision times were greater when judging genre-annotated summaries.

Table 6.12 Average summary decision times by task, genre annotation & rating (in seconds, excluding think-aloud task)

Summary Rating	Genre Annotation	Task 1	2	3	4	Average
1	With	12.1	**11.5**	**10.4**	13.5	12.0
	Without	9.5	14.9	11.1	13.3	12.1
2	With	**11.9**	16.9	12.3	17.9	15.1
	Without	15.7	13.9	11.9	12.2	13.5
3	With	12.3	**12.2**	11.2	14.5	12.6
	Without	11.5	12.7	11.1	12.1	11.9
4	With	9.8	11.1	**10.2**	16.3	12.0
	Without	8.5	10.5	11.0	10.3	10.1

Effects of Participants' Familiarity with the Search Topics

This previous topical search experience by some of the participants (reported near the beginning of the Results section) did not seem to affect the experimental results. Overall, these 15 participants performed 19 tasks (excluding the practice) in which they reported previous topical search experience. These comprise 14.8% of all tasks in this study (excluding practice). However, whatever experience they had did not manifest itself in more stable judgments, or faster decision times, of the summaries. Table 6.13 shows the stability of the judgments of the summaries for all 19 tasks: total, by presence of genre annotation, and by each of the four tasks. There do not seem to be any differences worth noting.

Table 6.13 Relevance decisions by participants in tasks with reported previous search experience for task topics

ALL	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	58	21	8	10	97	59.8%	0.69	0.69
2	42	26	18	16	102	25.5%	0.90	0.08
3	33	37	25	18	113	22.1%	1.07	(0.75)
4	11	20	12	24	67	35.8%	1.27	(1.27)
	144	104	63	68	379	35.1%	0.96	(0.25)

W/ GENRE	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	22	11	4	3	40	55.0%	0.70	0.70
2	19	16	11	9	55	29.1%	0.87	0.18
3	11	15	17	9	52	32.7%	0.88	(0.54)
4	8	11	7	6	32	18.8%	1.66	(1.66)
	60	53	39	27	179	34.1%	0.98	(0.24)

PLAIN	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	36	10	4	7	57	63.2%	0.68	0.68
2	23	10	7	7	47	21.3%	0.94	(0.04)
3	22	22	8	9	61	13.1%	1.23	(0.93)
4	3	9	5	18	35	51.4%	0.91	(0.91)
	84	51	24	41	200	36.0%	0.95	(0.26)

TASK 1	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	22	10	3	2	37	59.5%	0.59	0.59
2	12	8	7	6	33	24.2%	0.94	0.21
3	12	15	11	5	43	25.6%	1.02	(0.79)
4	7	10	5	4	26	15.4%	1.77	(1.77)
	53	43	26	17	139	32.4%	1.03	(0.37)

TASK 2	Full-text Rating							
Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	9	3	1	0	13	69.2%	0.38	0.38
2	10	6	0	3	19	31.6%	0.84	(0.21)
3	7	5	3	0	15	20.0%	1.27	(1.27)
4	1	3	4	5	13	38.5%	1.00	(1.00)
	27	17	8	8	60	38.3%	0.88	(0.52)

TASK 3 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	3	1	0	2	6	50.0%	1.17	1.17
2	0	0	0	1	1	0.0%	2.00	2.00
3	2	5	1	4	12	8.3%	1.08	(0.42)
4	0	0	1	0	1	0.0%	1.00	(1.00)
	5	6	2	7	20	20.0%	1.15	0.15

TASK 4 Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	24	7	4	6	41	58.5%	0.80	0.80
2	20	12	11	6	49	24.5%	0.88	0.06
3	12	12	10	9	43	23.3%	1.05	(0.63)
4	3	7	2	15	27	55.6%	0.93	(0.93)
	59	38	27	36	160	38.1%	0.91	(0.10)

Thirteen of the 19 tasks were tasks in which the summary decision times were not distorted by the think aloud procedure (performed as each participant's final task). For these tasks performed by participants' with reported topical search experience time, the average summary decision time was 11.4 seconds, compared with 12.5 seconds for all participants for all 128 tasks (which excluded the think aloud). It might seem that the previous topical search experience helped make faster summary decisions. However, the participants with some reported topical experience overall had an average decision time of 11.6, regardless of whether they reported topical experience with a task. It seems that this group of participants generally made faster summary decisions than the complete set of participants as a whole, thus, their decreased judgment time cannot be attributed to their reported experience.

Limitations

The purpose of this study was to see if the indication of a webpage's genre in the page's description would allow users to make faster or better decisions about the usefulness of the page for a given information need. If so, then web searching (and perhaps other types

of digital document retrieval) could be improved. To that end, the study had several limitations.

The circumstances under which participants were making these judgments differed from their usual searching experiences. First of all, the tasks were assigned. The information needs that they were looking to fulfill were not their own. As reported earlier, several participants mentioned this as something that made the process more difficult than what they considered to be typical searching. Specifically, they mentioned not being familiar enough with topics of some of the tasks. Given that topic is widely acknowledged to be the most important criterion for relevance, one who lacks familiarity with a topic might have to concentrate relatively more on topical considerations when making a usefulness judgment. Thus, someone concentrating on judging on whether a description was topically relevant could be less likely to take advantage of the non-topical clues that the genre of a document might offer. Of course, other variables, such as motivation, could also be expected to vary between assigned and real user information needs.

Another way in which the experimental procedure differed from participants' typical search experiences is that the tasks were not really searching tasks. Participants did not formulate queries, nor were any query keywords highlighted in the webpage descriptions. However, many participants reported using keywords in their decision-making, most likely because at least some of the keywords were fairly obvious, given the statements of information need (see Figures 6.1 and 6.2). Also, one might consider the omission of a typical relevance indicator (such as keyword highlighting) to be an advantage for detecting the impact of genre annotation, as the annotation would then have less competition for the user's attention.

The experiment differed from real-life searching in that search results were presented one at a time, and not ordered by relevance to the query. One participant mentioned that it was harder to make judgments about a single description when there were no others with which to compare it. Another difference between real-life and the experiment was that no pages were evaluated until after all the descriptions were evaluated. That actually might have helped equalize the description judgments as none were informed by information from viewing previous description/page combinations, which might have helped interpret succeeding descriptions.

Another difference from real searching was that participants had URL information during the description evaluation that was not available during the page evaluation (as the latter did not immediately follow the former). The URL could possibly contain additional information, or more quickly convey similar information for judging the page. Page evaluation was also different from real-life in that linked pages always opened in a separate window. This was to prevent participants from losing track of the original page that was being evaluated. A few participants had difficulty navigating back to the experimental window, and some even closed the experimental window in the process, and had to have the researcher restart the mechanism at the page which they were evaluating. (Fortunately, page evaluation times were not used in any of the analyses.) In any case, this awkwardness imposed by the experimental mechanism could conceivably have dissuaded participants from investigating as many links as they might, or may have affected their browsing behavior in some other way.

Discussion

Despite the aforementioned limitations, the overall indicativity for all decisions of 35.2% was consistent with that of the two known studies that have reported similar measures for webpage descriptions. In a study involving graduate student research, Lan (2002) reported indicativity of 41.3% (computed from a content analysis of subject's verbal judgments, without the use of a quantitative rating scale). Major differences in the experimental procedure were that Lan's subjects were motivated by their own real information needs, and that these needs were of a different nature than those of this study. Also, Lan's subjects participated in a think-aloud procedure during each search, and highlighted (on the screen) portions of the documents (and surrogates) that influenced each usefulness decision.

As reviewed in Chapter Two, one other study is known that quantitatively measured webpage description indicativity (Tuffs, 2002). Using data reported in this masters thesis, it was possible to calculate an overall indicativity of 56.5%, for ten subjects, each rating twenty pairs of webpages and descriptions, all performing the same assigned task. Like the current study, the complete set of descriptions was judged first, followed by the complete set of webpages. All documents were judged in order of relevance as provided by Google. Like the current study, participants were allowed to follow webpage links during the page evaluation process, but the exact details of how the experimental handled this, and the precise rating interface, are unclear from the paper. There was no report of linked pages opening in a separate window (and sometimes causing problems, as encountered here).

However, Tuff's (2002) study was not comparing different types of webpage descriptions. Instead, Google's descriptions were being compared to those of SpeechBot, a

spoken document retrieval system (http://speechbot.research.compaq.com/). A possible explanation for the higher reported indicativity is that the task approximates a known-item search (which includes the name of an album by a popular musical band). Figure 6.5 reproduces the task description. Thus, the task appears to be significantly less complex than those of the current study. Also, Tuff's study employed binary relevance judgments which also affects the indicativity measure. As shown in Table 6.2, collapsing the current study's four-point measurement scale to binary, in all possible combinations, results in an overall indicativity for this study, ranging from 60.9% to 72.3%, comparable to, if not better than, the 56.5% calculated from Tuffs' results.

Figure 6.5 Search task from (Tuffs, 2002, p.76)

> **Query:**
> Reviews of Oasis's new album Heathen Chemistry.
>
> **Scenario:**
> While listening to the radio one day you hear a song that you like by one of your favourite bands; Oasis. The DJ tells you that the song is taken from their recently released new album called Heathen Chemistry. You decide to find out about this new album by using Google to find reviews of it, to discover if it is as good as previous Oasis albums you already own and if it is worth buying.
>
> **Note:**
> Documents which are reviews of Oasis's Heathen Chemistry or documents that link to a review of the album are relevant. Documents reviewing previous Oasis albums or documents containing news about Oasis's latest releases are not relevant.

The few other studies that have measured the indicativity of document representations as compared to the actual documents involved collections of non-web documents. (These studies are summarized in Appendix A.) Saracevic (1969), using a collection of 600 documents on tropical diseases, reported indicativity (termed "immutability" in that paper)

of 85% for document titles, and 90% for abstracts. Using a collection of 20,000 documents "carefully selected…to serve an identifiable user population at MIT", Marcus, et al., (1978, p. 15), reported a range of indicativity from 64% to 73% for a variety of representations of varying lengths. Both studies (Saracevic, 1969; Marcus, et al., 1978) employed three-point relevance scales, and topical definitions of relevance. Both studies also employed intermediaries for finding the documents to be presented to the users. Kent, et al. (1967) used intermediaries to search document indices appropriate to the users, who were recruited from the medical complex at the University of Pittsburgh. For a variety of document representations of varying length, indicativity ranged from 69% to 82%, using a binary relevance scale. McLellan, et al., (2001) had participants search a collection of 1000 news articles to satisfy an assigned need on the topic of the history of the Russian-Chechen conflict. Using a 5-point relevance scale, indicativity ranges from 56% to 70% across three different, anonymous summarization systems. In both (Kent, et al., 1967) and (McLellan, et al., 2001), the relevance judgment was operationalized as whether or not the full-text of the document was worth obtaining. Although this collection of studies may be instructive for the design of document summary evaluation experiments, comparison of indicativity scores across studies is not useful because of the differences in summary types, document types, tasks, users, measurement scales, and composition of the collections. However, the size of the ranges of indicativity reported suggests that surrogate characteristics can affect the ability of users to predict document relevance, but (Marcus, et al., 1978) is the only study to actually report having measured statistically significant differences among the tested representation types.

Regarding the other quantitative measure of description decision time, only one other study measured it. Tuffs (2002) reported an average decision time of 10.3 seconds per web page descriptions by Google versus 12.5 seconds in this study. The longer decision time observed here is consistent with the assumption of the more complex nature of the participants' tasks.

Regarding participants' comments about web search engine description quality, sentiments expressed in the results here are echoed by both (Lan, 2002) and (Tuffs, 2002). Clearly, there is room for improvement in search engine web page descriptions. Says Lan, whose participants used eleven different web search engines:

> Participants did not like an excerpt taken out of context to be used as a surrogate. They did not want to have a summary which contained a few fragmented phrases or incomplete sentences. (p. 293)

Tuffs adds:

> Participants thought that the Google summaries appeared quite random and confusing in their selection of excerpts to form the summary and they are not flowing coherent pieces of text in their own right, this made judging relevance difficult. (p.58)

Overall, this study was unable to detect any quantitative evidence that genre-annotated descriptions had on the decision process. Neither indicativity nor decision time was improved, either overall or by task. It was a surprise to this researcher that only 17 of 32 participants reported even noticing the genre label in the descriptions! Even just looking at the data (not reported in the Results) from those seventeen, the pattern of results was unchanged. It didn't even seem to matter if the participants were aware of the genre label.

What is puzzling is that 12 of the 17 who reported being aware of the label stated that they thought the label was helpful. Comments from participants are consistent with those reported by Lan (2002), in which document type/genre was the second most mentioned class

of relevance clue (305 mentions, 16% of all document characteristic mentions), second only to topic. All thirty-eight participants mentioned it, noting 49 different document types.

Why didn't the perceived helpfulness reported in this study translate into better measurable performance? Any of a number of reasons is possible, having to do with any or all of the participants, the tasks, the pages selected, the genres, or other aspects of the experimental procedure. The variability of these factors will be discussed in the next chapter.

One factor that might have had a possible "scrambling" effect on the relevance judgments was the length of the experiment, which averaged about 105 minutes, and sometimes went as long as 2.5 hours. (Note that after pilot testing with experiment with two participants, the original design was changed by reducing the number of pages in the practice task from 20 to 5, and by limiting the think-aloud to just the first ten summaries and the first ten web pages in the final task.)

Even with these modifications, some participants may have felt time pressure (and possibly fatigue), which may have affected their judgments. Possible ways to reduce potential time pressure include: reducing the number of tasks, eliminating the think-aloud procedure, and disallowing the following of links during the page evaluation.

Certainly, having more subjects do fewer tasks each could maintain the same quantity of data, while reducing possible time pressure. Even one less task would be helpful. Regarding the think aloud procedure, it would seem that there is now enough qualitative data regarding the users' perceptions of genres in web searching. This modification could save 10-15 minutes per subject. Disallowing link following during the page evaluation would be a less favored approach. Following links is normal searching behavior and should theoretically improve the accuracy of the page judgments.

Another modification would be to tell everyone about the genre labels at the beginning, and even to explain a little about genre and why it might be helpful. It is likely that users are not accustomed to seeing search results labeled with genre, and they may need to be introduced to it. Also, now that there is qualitative evidence that many perceive genre labeling to be helpful, there is less need to be concerned that the instruction in genre will bias the participants.

It was the researcher's impression that participants' confidence in their judgments of the descriptions tended to vary by individual description. Perhaps adding a confidence rating to each usefulness judgment would add a piece of data that would be useful for understanding what is happening. For example, do judgments change more when confidence is lower? Could genre labeling instill a sense of false confidence in the page usefulness rating? Some participants in this study expressed surprise when summaries they thought would not be useful (either because of the perceived genre or the genre label) turned out to be very useful after all. Of course, this additional piece of data would add to the length of the session.

One drawback to the experimental design here, as compared to the other surrogate evaluation studies, is that participants in this study did not rate both types of surrogates for a single task. They either rated a genre-labeled description and the corresponding page, or they rated a non-genre labeled description and the page, but never both types of descriptions. A design that had participants rate both types of surrogates, either successively (as in Janes (1991b) and Marcus, et al. (1978)), or as part of a set of surrogates (all non-genre, followed by all genre, both sets randomly order) could pinpoint changes specifically to the genre label, and provide greater statistical control over subject variance. A possible problem is that

subjects' first judgment could influence the second one. Some surrogates could be presented twice to get a sense of the consistency of each subject's ratings.

In summary, qualitative evidence from multiple studies shows that users think that document genre is helpful for determining usefulness. A future analysis of the think-aloud data collected here may provide more specific insights into searchers' use of genre in the surrogate evaluation process. Other additional studies can focus on measuring performance data in order to show that genre labeling can make a statistically significant difference.

In the next chapter, this study is discussed further in the context of the complete set of studies conducted for this dissertation, and in the context of the overall use of genre in improving the entire web search process.

This page intentionally left blank.

CHAPTER 7

DISCUSSION AND SUMMARY

With the advent of large, digital collections of heterogenous documents (namely the World Wide Web, and large sub-collections thereof), attention to the prospect of identifying web genres in order to enhance searching has recently increased (e.g., Kennedy & Shepherd, 2005; Santini, 2005; Crowston & Kwasnik, 2004; Stein & Zu Eissen, 2004).

Potential uses of genre include annotation of search results (as investigated in this dissertation), constraining search results, implicit or explicit relevance feedback, and as an aid to visualizing collection contents prior to query formulation. Crowston and Kwasnik (2004) also suggest that search results might be tailored by document genre. It is possible that additional uses may be proposed in the future. For example, Rose & Levinson (2004) have begun development of a web query typology, in which some query types whose results may closely coincide with some web genres.

As explained in Chapter One, genre is essentially a context document descriptor, representing document context rather than content. In Chapter One, three criteria of context descriptors for the improvement of document representations in IR systems were proposed: the degree of the descriptor's "public-ness" or "private-ness", the usefulness of the descriptor for the search tasks undertaken by the system's users, and the ability to classify documents according to the descriptor algorithmically.

This dissertation explored the viability of genre as a document descriptor according to the first two criteria. The work has contributed to the body of knowledge about web genre in several ways. First, it has been shown that a genre palette can be constructed for the edu domain in which users can generally agree on the genres of individual web pages. Certainly, some refinements can be made, and those have been suggested in Chapter Five. Thus, it seems that at least some genres have an adequate amount of "public-ness" for use as a context descriptor. Secondly, participants who were aware of the genre annotation in the search results responded positively to the feature, and reported it having been useful to them in evaluating search results. This is consistent with observations from, and user comments solicited during, Lan's qualitative web searching study (Lan, 2002). However, this study has been unable to show quantitatively that the genre annotation improved the consistency of user judgments (as compared to their judgments of the corresponding webpages), and unable to show that users were able to make faster judgments when the genre annotation was present. Yet, participants' positive attitudes toward genre certainly encourage future work in this area. Nevertheless, there is still no hard evidence to show that automated genre identification would make a difference, and if so, under what circumstances. In this arena, therefore, a contribution of this dissertation is to delineate the problem of adding genre annotations to information systems retrieval in more detail, and to discuss some of the difficulties inherent in designing research to investigate the idea.

One speculation is that one of the major difficulties in showing that the use of genre in the search process makes an empirically measurable difference is that the added value of using genre varies with primary search process variables: the users, the tasks, the collection,

and the surrogates describing the documents in the collection. Each of these variables is now discussed in turn.

Issues in Experimental Design

Users

As mentioned earlier, since the user group (aka the discourse community) is part of the definition of a genre, the value of a specific genre depends heavily on whether the user understands the genres of the documents in the collection. In measuring the value of genre for retrieval this factor can be addressed by using a palette of broadly-defined genres, such that most if not all of the users will know the genres. However, just as the value of genre is unclear, it is also unclear how much value resides in broader genres, and how they interact with the other search process variables such the tasks and collection. It is not suspected that genre recognition was a problem for the current study in that the users seemed to understand the genre palette. However, there may have been some variability in terms of how much users tried to infer genre from the unlabeled document summaries. As mentioned earlier, some short user training with the genre label before the actual tasks might have made the genre annotation more useful.

Tasks

Although there is no evidence to prove this, it seems like a fairly commonsense notion that (in most cases) all of the genres in a palette would not be useful to most tasks. A genre is useful to a search task for one of two reasons. One, the chance of a document of a specific genre being relevant is greater than the average chance of a document in the collection being relevant. For example, the “FAQ/Help” web genre might be more applicable to the information need concerning “how to get rid of ants in the home” than a document of

some random genre might be. The other reason a genre is useful to a search task is when a document of this genre is most likely or certainly not relevant, and the chances of that genre (due to the interaction between the search terms and the collection) will turn up in the search results in non-trivial amounts is high. For example, with a need to find information on children's merchandise in the "Strawberry Shortcake" collection, it might benefit the searcher to exclude documents from the "recipes" genre. Thus, certain genres are only useful to certain tasks. However, as alluded to above, the value of genre may vary depending on the specific terms in the query. Even for the same task, the amount of helpfulness added by genre may vary.

It is possible that, in general, a majority of tasks could be orthogonal to a majority of genres. Berkenkotter and Huckin (1993) point out that "recurring situations resemble each other only in certain ways and only to a certain degree" and that "socially induced perceptions of commonality do not eradicate subjective perceptions of difference" (p. 481). In other words, the genre of a document may not be a strong indicator of relevance. Documents and their possible interpretations are multi-dimensional. One can imagine both relevant and irrelevant documents of the same genre for a given query. In a given search situation, it is may be unclear whether genres are orthogonal to the particular information need.

The Collection

As noted by Blair (2002a), the size of the collection influences the difficulty of the retrieval task. Also, the relative distribution of genres throughout the collection affects the individual genre's value for constraining searches. The content of the collection's documents

would also play a role in affecting whether a genre's effect on search constraint is useful or not.

Webpage Surrogates

A participant in this dissertation's study involving genre-annotated surrogates remarked that oftentimes the summary already contains genre information. Genre information can be inferred from titles, URLs, excerpts or any combination thereof. Thus, the specific content of surrogates will affect the value of including explicitly labeled genre in improving retrieval. In search engines that use query-biased summaries (e.g., Google; see also White, et al., 2003a), then the user's search terms affect the content of the summary which may in turn affect how much extra information a user can infer from a genre label.

Given the difficulty in demonstrating the benefit of using web genre to improve the search process, it may be asked whether it is really worth the effort, especially in light of the negative arguments that can be made regarding genre's potential value. For example, it is widely expressed that genres are dynamic and change over time (e.g., Kennedy & Shepherd, 2005). Thus, even if the use of genre is found to provide some benefit, whatever efforts were taken to establish the genre palette and develop the algorithms for automatic classification, these efforts will need to be repeated as the palette changes. That also creates the additional question, how does one know when it is time to recalibrate the genre palette?

There are more arguments against the use of genre. As mentioned earlier, many web summaries already indicate genre in some way, making an annotation unnecessary in those cases. Also, because web pages are hyperlinked, the page of an unwanted genre may link to a page of a desired genre, thus possibly causing valuable pages to be overlooked. That

phenomenon was encountered by participants in the study of genre annotated search results (Chapter 6) when they were surprised to find that a personal website page or a blog linked to information they were looking for.

One participant in the study, who reported liking the genre labeling, said that the more specific genres (like FAQ/Help) were more useful than vague genres (like article or welcome/homepage). Unfortunately, as Table 6.1 demonstrates, it seems that a high percentage of webpages belong to those "vague" genres. Unless these genres (possibly including index/table of contents/links as another common but vague genre) can be more clearly defined or refined, much of the potential of genre to improve search may be lost (at least in general collections).

Also, in some circumstances, it may be possible and productive to constrain by genre using keywords, e.g., the use of "review" when looking for a review of a product or an artistic work. In other cases, the user may not know enough about what exists in the collection to know what genre(s) to use to constrain a search. For example, if someone is seeking to compare computer science lab facilities at other schools, what is the "best" goal? Should it be a "list of links" to computer science departments? Should it be a set of search results containing computer science department "homepages"? Would it be better to look for the "article" genre, in hopes of finding pages with lots of descriptive text? It seems hard to know which is best, if one has not already done the search!

Another large negative for investing time in trying to demonstrate that genre is useful for web search, is that regardless of how useful it is shown to be, nothing can be implemented unless classifiers for automatic genre identification are/can be developed. Web page classification still seems to be in a developmental phase (Rosso, 2003). Studies which

have classified web pages into genres have reported only marginal results at best (Karlgren, et al., 1998; Lee & Myaeng, 2002, Stein & Zu Eissen, 2004; Kennedy & Shepherd, 2005). Another classification effort which was reported to be "underway" (Roussinov, et al., 2001) has not since been updated in the literature. (It should be noted that machine classification cannot be expected to be more accurate than human classification, and the level of accuracy that would be necessary for genre to be useful, is unclear.) Even assuming that the classification accuracy is acceptable, developing the classifiers is a non-trivial task, and one that would have to be revisited, to some extent, each time it is decided to update the genre palette.

If researchers fail to demonstrate circumstances under which the use of genre would benefit web search, given the positive responses of experimental participants, one could proceed to implement a system, and then attempt to demonstrate the benefits. Assuming the automatic classification was successful, an actual running system would provide an advantage in that it would be possible to observe realistic searches with users motivated to find information important to them, while forming their own queries, and judging search results and webpages in a more realistic environment.

One approach might be to start automatic classification efforts on those well-recognized genres that have the most characteristic form. One would expect those genres to be the easiest for which to construct an automatic classifier. For example, FAQs and discussion archives seem to have prototypical forms familiar to web users. Although the results here indicate that the personal website and form genres will need some refinements, these genres, too. might easily be amenable to automatic classification with relatively less

effort than many of the other genres, e.g., the article genre which has much less distinctive form (as well as other "problems" as discussed in Chapter Five).

Regardless of the path taken to construct a live system, the use of genre would have to provide a benefit greater than the value of the extra storage and processing time needed to implement this extra feature. Each time a document is indexed, extra processing time would be needed to determine the genre. Extra storage to store a genre ID per page would be trivial. Pages of search results containing 10 summaries might be lengthened by 10 lines, depending on the display format. So, classification time would seem like the major cost. However, assuming that, for a search engine like Google, there are always more pages to index and re-index than could possibly be completed, any feature which requires processing time necessarily reduces the size or the freshness of the index (given the same amount of computing resources). Thus, the value of the feature must be greater than the cost of the reduced size or freshness of the search engine's webpage index.

The Next Step

It seems that there is just as much anecdotal evidence against the prospect of using genre to improve web search as there is for it. However, people's reactions to the idea are generally positive, providing some additional motivation to continue the research. There are three alternatives for going forward from here. If the genre effort is to continue, the next step would be to refine and redefine the genres in the current genre palette. This could be accomplished with varying user groups and varying sets of web pages in order to confirm and refine the palette. At this point, the work could branch into one or both of the following two alternatives. In alternative one, researchers could continue to attempt to demonstrate the

value of genre for information retrieval. Better performance might be obtained by virtue of using a refined palette. A combination of other users, other tasks, other pages, etc. might also make a difference. Other ideas for improving the experimental design were covered in Chapter Six. In alternative two, after refining the palette, work could proceed straight to the machine classification of webpages into genres. Finally, in alternative three, the efforts to implement webpage description could be abandoned, and research efforts could be applied to other search enhancements that seem more likely to succeed.

Given the number of variables that affect the usefulness of genre, it is doubtful that experimentation with artificial tasks, even with a variety of users, tasks and collections, could give an realistic picture of genre's usefulness (or lack thereof). The path of alternative one could be a long one, and really not prove much in the end. Much more realistic experimentation under a variety of conditions could be accomplished easily, if the web genres could be automatically assigned to the pages. Of course, that is a BIG if. Thus, it seems like the choice of future direction is between proceeding with automatic classification efforts (after palette refinement), or abandoning the idea in favor of potential search improvement ideas.

Alternative one could be attempted one more time in an attempt to provide a proof of concept, using refinements to the experimental procedure suggested in Chapter Six. If results were not promising, and did not suggest at least suggest task/genre combinations worth exploring, then a decision would have to be made: either commit to the resources necessary to explore machine classification or abandon the effort all together.

Or possibly, the edu domain is still too broad to allow us to get a handle on the problem. Given the strong overlap between this palette and the internet-wide palettes of

others (see Table 5.10), maybe the edu domain is still too much like the whole web, which as a basis for defining a user group has been deemed "both conceptually weak and pragmatically untenable" (Nilan, et al., 2001, p. 335).

Still, the authors continued:

> We will narrow our search and try again…We believe that this study will provide the insight needed by automated systems designers to reliably re-create the classification of Web pages. Further, we believe that this study will enable the designers to represent search results to users by employing terms…that are inherently meaningful to users. (p. 336)

Searching by genre is an appealing idea, yet a difficult one in which to figure out exactly how to implement it. Like the general information retrieval problem, it is a hard one. Of course, that means that researchers will be afforded many opportunities to pursue their ideas toward making it work.

APPENDIX A

PAST STUDIES COMPARING RELEVANCE ASSESSMENTS OF DOCUMENT SURROGATES VERSUS FULL-TEXT

	Kent, et al. (1967)	Saracevic (1969)	Marcus, et al. (1978)	Tombros (1997)	Mani, et al. (1999)	McLellan, et al. (2001)
# of tasks per subject	1	Variable	1	5	20	1
Source of task problem	User need	User need	User need	TREC topic	TREC topic	Simulated work task situation
Judgment scale	Binary	3-point	3-point	Binary	Binary	5-point
# of judgments per document	2	3	5	1	1	2
Types of objects judged	First: either Citation, Citation + Abstract, Citation + 1st par., Citation + last par., or Citation + 1st and last pars. Second; full-text	Title, Abstract, Full-text In that order	First four (in random order): Title, Abstract, Subject terms, Subject terms that (partially) match query terms Last: full-text	Summary and full-text with no prescribed pattern of exposure (determined by "search" behavior of subject)	One of three different types of summaries and full-text	Summary, Full-text
Documents per task	10-25	Variable	20	50	50	16
Ordering of document judgments	Surrogates randomized, then all full-text randomized	By document (unspecified order of documents)	By document (randomized document order)	By ranked list of documents	By ranked list of documents	Surrogates randomized, then all full-text randomized

APPENDIX B

INSTRUCTIONS FOR PARTICIPANTS OF CARD SORTING STUDY

On the table in front of you is a stack of printouts of 100 web pages from web sites in the Internet ".edu" domain. This exercise involves deciding how to separate the printouts into piles. Your task is to place web pages of the same "genre" in the same pile. You may create as many piles as you wish. A pile may have as many or as few printouts as you think belong together.

A document genre is a category of documents characterized by similarity of function, style, form or content. Traditional document genres include, for example, business letters, cooking recipes and greeting cards. Note that a document's genre is not the same as its subject. For example, a business letter may be about the availability of a new product, or an invitation to interview for a job. Both examples are business letters, but their subjects are different.

Each pile you create should represent a different genre from the Internet .edu domain of web pages. Sometime during the exercise, you should come up with a name for the genre of each pile. Post-it notes are available for labeling each pile with its name. When you are finished with making the piles and naming them, I will ask you questions about your pile names, and the printouts in each pile.

If you have any questions or comments now, or during the exercise, please let me know.

Thank you for your help.

APPENDIX C

LIST OF 48 GENRES FOR USE BY PARTICIPANTS IN GENRE REFINEMENT STUDY

	genre	description
A1.	about	short description of purpose or objectives of an organization
A2.	abstract	title and brief description and reference (one page)
A3.	advice	advice on how to deal with a situation
A4.	article –1	something about a topic with supporting facts or opinions; tells a story somehow; not conversational
A5.	article –2	several pages talking about something (not as time sensitive as enews; more topic focused rather than event focused
B1.	bibliography	page of pointers to books or articles; lists of pointers to papers, books, or other resources (that are usually on other sites)
B2.	biography	page primarily about a person
B3.	blurb	description page for a place or program (used to find out if you want to go there)
C1.	card catalog	reference to a titled work
C2.	contact form	for asking a question
C3.	conversations, observations, or opinions	opinions, stream of consciousness stuff, just talking
C4.	course	classes or programs offered for instructional purposes
C5.	course description	what's covered in a course; syllabus
C6.	course list	page that lists courses
D1.	database	for access to a database (a search engine)
D2.	diaries, weblogs or blogs	a personal narrative or time log of activities (not a biography)
E1.	email	form for sending email
E2.	enews	online articles
F1.	FAQ	frequently asked questions; questions may be links to answers; not interactive like a forum
F2.	form	page for entering info
F3.	forum/interactive discussion archive	one or more messages and/or responses that are viewable by an audience
F4.	full-text index	page pointing to full-text of a book or article or magazine or journal
H1.	help	assisting you to perform a task (like an FAQ, but links are topics, not questions)
H2.	history	like "reference" but about the past
H3.	homepage	mission statement and table of contents for an organization
I1.	indices/table of contents/links	page which is primarily a list of links
I2.	instructional	recipe to follow a task
J1.	job listing	describes one or more jobs that are available
J2.	joke	story intended to be funny
N1.	navigation	top-level pages with list of links (to same site)
N2.	news index	headers for online articles (enews postings)
N3.	newsletter	fairly current news for an organization or a group

(continued)

APPENDIX C

LIST OF 48 GENRES FOR USE BY PARTICIPANTS (CONTINUED)

P1.	personal website	page that somebody writes about themselves
P2.	picture/photo	page primarily containing a picture with few or no words
P3.	poetry	contains a poetry or similar wordplay
P4.	product for sale	page from an online store
P5.	program description	describing educational programs
P6.	publications, bulletins, newletters	information published about/by various organizations; collection of articles (or links to articles)
R1.	reference	detailed facts about a subject
R2.	registration	form for registration
R3.	resume	for looking for a job
R4.	review	short description of literature, art, TV, etc.
S1.	search start	place to type-in key words and search
S2.	shopping	for purchasing products
S3.	speech	text of a speech
S4.	story	shorter prose than reference, complete in itself, fiction or non-fiction
S5.	syllabus	course syllabus
W1.	welcome page	starting page (does not have to be the "top" page in a site)
XX.	SUGGEST YOUR OWN	

APPENDIX D

INSTRUCTIONS FOR PARTICIPANTS OF GENRE REFINEMENT STUDY

This exercise is about deciding what "genre" a web page belongs to. Following is the definition of genre that we will be using.

A document genre is a category of documents characterized by similarity of function, style, form or content. Traditional document genres include, for example, business letters, cooking recipes and greeting cards. Note that a document's genre is not the same as its subject. For example, a business letter may be about the availability of a new product, or an invitation to interview for a job. Both examples are business letters, but their subjects are different.

You have been given a list of 48 genre names and definitions that have been collected from people in a previous study. All of the genres describe web pages from web sites in the Internet ".edu" domain. Some of them may seem very similar, with only minor differences.

On the table in front of you is a stack of printouts of approximately 100 web pages from web sites in the Internet ".edu" domain. For each web page, choose the genre from the list which best describes the page, and write its number on the answer sheet in the space provided. If none of the genres seem appropriate, you can suggest your own genre name and definition, by writing them on the answer sheet in the space provided.

If you have any questions or comments now, or during the exercise, please let me know.

Thank you for your help.

APPENDIX E

ACADEMIC GROUPING – GENRE REFINEMENT STUDY

Page #	Votes by Page for each of the Five Genres					Total
	C4	C5	C6	P5	S5	
17	1.5		8	0.5		10
21	2		8			10
23		5.5			4.5	10
35		7			3	10
49	2		3.5	0.5		6
52		0.5		5		5.5
91		4				4
125			1			1
126		1				1
142			1			1
157				3		3
190		2.5		3.5		6

Genre names from Appendix C (see for genre descriptions)
C4 – course
C5 – course description
C6 – course list
P5 – program description
S5 – syllabus

APPENDIX F

PERSONAL GROUPING – GENRE REFINEMENT STUDY

	Votes by Page for each of the Four Genres				
Page #	B2	D2	P1	R3	Total
57			2.5		2.5
59	1				1
60	2				2
86	1	0.5	1.5		3
87	1	4.5			5.5
120		1			1
130		7	2		9
140		1.5			1.5
148	8				8
150				0.5	0.5
152		9			9
167	1				1
169			2		2
171		1			1
172		2	1		3
201			5.5		5.5
202	2.5		4.5	2	9
203			9.5		9.5

Genre names from Appendix C (see for genre descriptions)
B2 – biography
D2 – diaries, weblogs or blogs
P1 – personal website
R3 – resume
S5 – syllabus

APPENDIX G

QUESTION & ANSWER GROUPING – GENRE REFINEMENT STUDY

	Votes by Page for each of the Four Genres				
Page #	C3	F1	F3	H1	Total
6				1	1
16		3.5		5	8.5
37	1				1
40	1				1
43			2		2
57				1	1
59	2				2
86	6				6
87	0.5				0.5
88	1				1
109	0.5				0.5
120	3		4		7
128	2	1	1		4
130			1		1
132		0.5		1	1.5
140	1				1
143				1	1
144				1	1
149				2	2
151		4		1	5
154	1	1	8		10
171	1		8		9
172	3		4		7
177	1.5		8.5		10
178		3	6		9
179		10			10
197		1.5		3	4.5
200	2	1	5		8

Genre names from Appendix C (see for genre descriptions)
C3 – conversations, observations or opinions
F1 - FAQ
F3 – forum/interactive discussion archive
H1 – Help

APPENDIX H

TEXT GROUPING – GENRE REFINEMENT STUDY

Votes by Page for each of the Sixteen Genres																	
Page #	A2	A3	A4	A5	B3	E2	H2	I2	J1	J2	N3	P6	R1	R4	S3	S4	Total
4												1					1
6											1						1
9													0.33				0.33
11												2					2
22								3									3
26						1					1						2
27	3				1												4
34											1	1					2
37			3.5	1		3										1.5	9
40			4	2.33		2.66											8.99
42						1	1				1	2					5
47								5					1				6
52					1												1
57					1												1
58			5	1											3		9
59										5						1	6
60			2	3		1					1						7
86										1							1
87				1												3	4
88		7.5		1							0.5						9
91	3		2	1													6
96					1				9								10
98			1.5			6.5					1					1	10
99																1	1
109															9.5		9.5
110														1			1
111			3.5	1		5.5											10
112		4		4									1.5				9.5
118							1										1
120						1				1							2
125												1					1
126			1	3				2.5					1.5		1		9
127	1		1	5									2.5				9.5
128			4				1							1			6
132		3						4									7
134			1			2								7			10
135			2	2			4.5						1.5				10
136			5	4			1										10

APPENDIX H

TEXT GROUPING – GENRE REFINEMENT STUDY

Votes by Page for each of the Sixteen Genres																	
Page #	A2	A3	A4	A5	B3	E2	H2	I2	J1	J2	N3	P6	R1	R4	S3	S4	Total
140				1												5	6
142									9								9
143			1	1	1									1			4
144		5.5						1					0.33				6.83
148						1					1						2
150		3.5	0.5	3.5									1.5				9
151					1								1				2
152																1	1
157		1															1
163	1.5													7.5			9
164					0.5												0.5
169					2												2
170					1												1
174		3.5	2	2.5		1						1					10
176		6		2.5				0.5				0.5					9.5
178													1				1
190	1				1						2						4
193					1						2	2					5
194				8									2				10
198		8	0.33					1.33					0.33				9.99
200																1	1

Note: table continued from previous page.

Genre names from Appendix C (see for genre descriptions)

A2 – abstract
A3 – advice
A4 – article-1
A5 – article-2
B3 – blurb
E2 – enews
H2 – history
R4 – review
S3 – speech
S4 - story
I2 – instructional
J1– joke
J2 – job listing
N3 – newsletter
P6 – publications, bulletins, newsletters
R1 – reference

APPENDIX I

POINTERS GROUPING – GENRE REFINEMENT STUDY

	Votes by Page for each of the Eight Genres								
Page #	B1	C1	F4	H3	I1	N1	N2	W1	Total
1	1			1		2		6	10
2			3		3	1.5		2.5	10
3				8				2	10
4	1		2	1	1	1		3	9
5				7.5		0.5		2	10
6				1	2.5	2.5		1	7
8				3.83		0.33	4	1.5	9.66
9			1		7.83	0.33			9.16
10				3.5		0.5		6	10
11				2.83	3.33	1		0.83	7.99
16					1	0.5			1.5
20				1	3.83	2.83		2.33	9.99
22					2			2	4
24				5		2		3	10
26					2		6		8
27		6							6
28					9	1			10
34				2	2	2		1	7
36					10				10
42					2		3		5
43					8				8
49	2				2				4
57	0.5				3				3.5
58								1	1
118		1	3	0.5	1	1		1	7.5
125	2				4	1	1		8
129				2		2		4.5	8.5
132								1	1
137		1	6		1			1	9
143				1	0.5	1		2.5	5
144					0.33	0.5		0.33	1.16
149	1				4.5	2			7.5
151						1			1
157				2				2	4
162	1	3	4.5					0.5	9
164				3					3
167	4	1	1		2				8
170						1		2	3

APPENDIX I

POINTERS GROUPING – GENRE REFINEMENT STUDY

	Votes by Page for each of the Eight Genres								
Page #	B1	C1	F4	H3	I1	N1	N2	W1	Total
201	1				3			0.5	4.5
202								0.5	0.5
203					0.5				0.5
207	0.5								0.5

Genre names from Appendix C (see for genre descriptions)
B1 – bibliography
C1 – card catalog
F4 – full-text index
H3 – homepage
I1 – index/table of contents/links
N1 – navigation
N2 – news index
W1 – welcome page

APPENDIX J

FORM GROUPING – GENRE REFINEMENT STUDY

	Votes by Page for each of the Four Genres				
Page #	C2	E1	F2	R2	Total
22				1	1
47			1		1
197	1				1
204	4	1	4	1	10
206	1.5	5	3.5		10

Genre names from Appendix C (see for genre descriptions)
C2 – contact form
E1 – email
F2 – form
R2 – registration

APPENDIX K

SEARCH GROUPING – GENRE REFINEMENT STUDY

Votes by Page for each of the Two Genres			
Page #	D1	S1	Total
6	1		1
129	1		1
207	5	4.5	9.5
208	5	5	10

Genre names from Appendix C (see for genre descriptions)
D1 – database
S1 – search start

APPENDIX L

SALES GROUPING – GENRE REFINEMENT STUDY

Votes by Page for each of the Two Genres			
Page #	P4	S2	Total
110	8	1	9
163	0.5	0.5	1
165	8	2	10

Genre names from Appendix C (see for genre descriptions)
P4 – product for sale
S2 – shopping

APPENDIX M

MISCELLANEOUS GROUPING – GENRE REFINEMENT STUDY

	Votes By Page for The Other Genres			
Page #	A1	P2	P3	XX
8				0.33
9				0.5
18		10		
22	2			
34				1
47				3
52	3			0.5
57				2
59				1
60	1			
99			9	
112				0.5
118	1			0.5
130				
132				0.5
137				1
140	0.5			1
144				1
146			9	1
149				0.5
150				0.5
151	2			
157	2			
161			9	1
162				1
164	6.5			
167				1
169		6		
170				6
176				0.5
193				1
200				1
202				0.5

Genre names from Appendix C (see for genre descriptions)
A1 – about
P2 – picture/photo
P3 – poetry
XX – SUGGEST YOUR OWN

APPENDIX N

NEW PALETTE OF 18 GENRES

	genre	description
1	article	something about a topic, often with supporting facts or opinions
2	course description	what's covered in a course; syllabus
3	course list	page that lists courses
4	diary, weblog or blog	a personal narrative or time log of activities (not a biographical article)
5	FAQ/Help	frequently asked questions, or assistance in helping you perform a task; questions may be links to answers, or topics may be links to assistance; not interactive like a forum
6	form	page primarily for entering and submitting information (other than a search engine)
7	forum/interactive discussion archive	one or more messages and/or responses that are viewable by an audience
8	index/table of contents/links	a page which is primarily a list of links or text items ordered (usually alphabetically) so that a list item can be found easily, AND the page does not belong to any of the other categories
9	job listing	describes one or more jobs that are available
10	other instructional materials	materials (other than a syllabus) used in teaching course, including but not limited to tests, quizzes, assignments, answer keys, etc.
11	personal website	page (possibly a homepage) that somebody writes about oneself (but not a biographical article)
12	picture/photo	page primarily containing a picture or pictures with few or no words (other than captions)
13	poetry	contains poetry or similar wordplay
14	product for sale/shopping	for purchasing products (not a product review article)
15	search start	page primarily to enter key words and search a database; a search engine
16	speech	text of a speech
17	welcome/homepage	starting page (does not have to be the "top" page in a site); may contain introductory information about a specific organization, department, program, etc. and a table of contents
18	NONE OF THE ABOVE	page that definitely does not fit into any of the above categories

APPENDIX O

OPENING PAGE OF THE GENRE PALETTE VALIDATION STUDY

Categorization of Higher Education Web Pages

Purpose of this Study

We are inviting you to take part in research designed to help further the development of new techniques for searching the World Wide Web. In particular, we are studying how web page categories can be used in search engines to help searchers find what they are looking for more quickly. This specific project focuses on people's abilities to categorize WWW pages of higher education institutions. The Principal Investigator of this project is Mark Rosso (advised by Dr. Stephanie Haas).

What Will Happen During the Study

Roughly 30 people (18 years of age or older) will participate in the study. As a participant, we will ask you to perform three tasks. First, you will be asked to indicate that you understand your rights as a participant. Next, you will be asked to answer five anonymous demographic questions about yourself. Finally, for each of 55 web pages, you will be asked to look at the web pages one at a time, and to pick the best category (or categories) to which it should be assigned. (The survey works best using Internet Explorer 6.x as your web browser.)

If you have any questions regarding this study, please contact Mark Rosso (919-489-7969, rossm@ils.unc.edu). Or you may contact his faculty advisor, Stephanie Haas, (919-962-8360, stephani@ils.unc.edu).

Your Privacy is Important

We will make every effort to protect your privacy. All information recorded by this web form will be kept anonymous. Since we will be making efforts to protect your privacy, we ask you to agree that we may use any information we get from this research study in any way we think is best for publication or education.

Risks and Discomforts

We do not know of any personal risk or discomfort you will have from being in this study.

Your Rights

Your participation in the study will last approximately 30 to 45 minutes. You are free to refuse to participate or to withdraw from the study at any time without penalty and without jeopardy.

Institutional Review Board Approval

The Academic Affairs Institutional Review Board (AA-IRB) of the University of North Carolina at Chapel Hill has approved this study. If you have any concerns about your rights as a participant in this study, you may contact the AA-IRB at 919-962-7761 or at aa-irb@unc.edu.

I have had the chance to ask any questions I have about this study, and they have been answered for me.
I have read the information in this consent form, and I agree to be in the study.

I agree to these Terms

APPENDIX P

DEMOGRAPHICS/INSTRUCTIONS PAGE
GENRE PALETTE VALIDATION STUDY

Welcome to the web page categorization study

Please fill in the following information, and read the following directions carefully.

Your age (in years): [ ]

Your highest grade level of education completed: [—]

Are you currently enrolled as a full-time student? ○ yes ○ no

Your estimate of the number hours per week on average that you spend using the World Wide Web: [ ]

Your estimate of the number hours per week on average
that you spend searching the World Wide Web for information: [ ]

Directions

You are about to be shown a series of 55 web pages from the edu (educational) domain of the Internet.

For each web page you are shown, at the bottom of your screen, you will see a menu box with the names of various web page "genres". You are to decide which genre best describes this page. Then, with your mouse, please choose that genre from the menu box. If none are appropriate, you may decide "NONE OF THE ABOVE". For an explanation of the genres, you may click on the "Web Page Genre Definitions" button at the bottom of the screen, at any time.

Do not at any time use the "Back" capability of your browser to view previous pages, and do not click on any links in the pages you are categorizing. Also, if you close your browser before finishing, you will have to start the survey from the beginning in order for your responses to be counted in the experiment. We appreciate your participation.

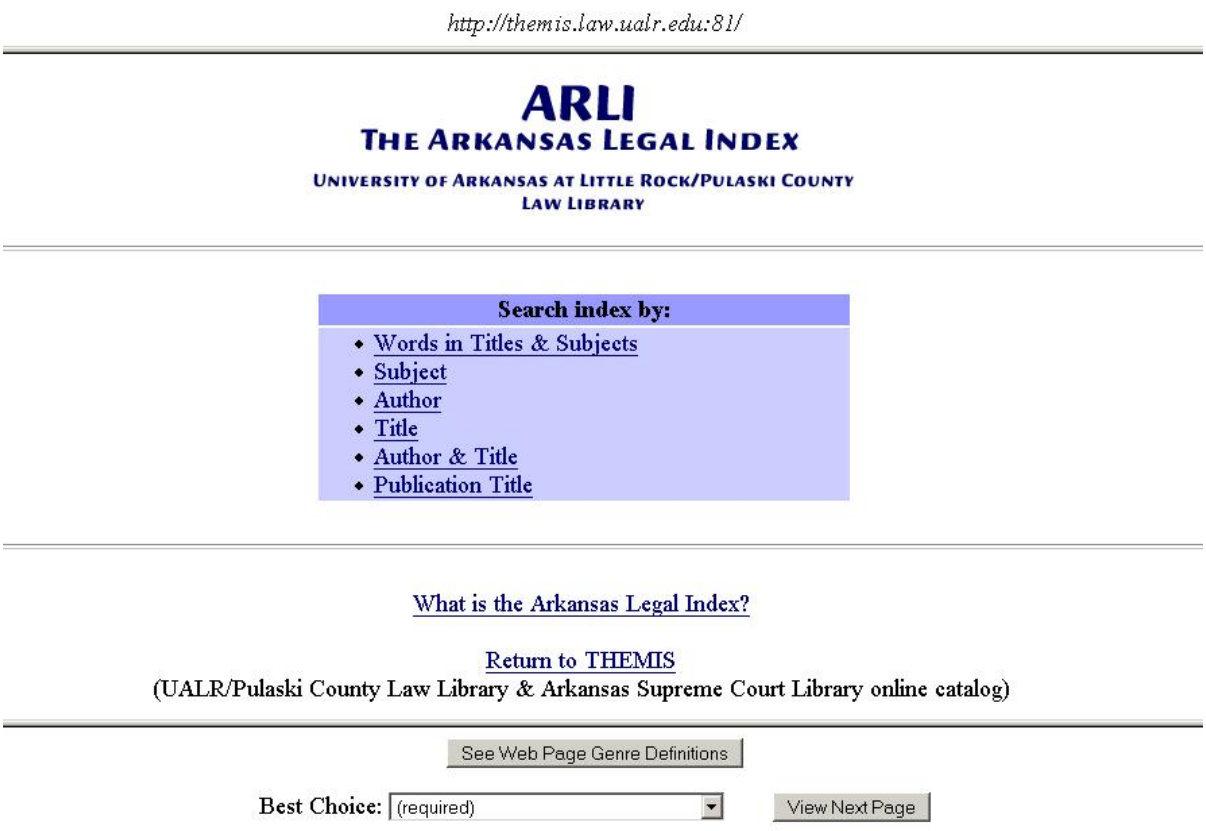
Please click the "Web Page Genre Definitions" button now to read an explanation of the genres. Then click the following button to begin viewing your first web page.

[Web Page Genre Definitions]

[Begin]

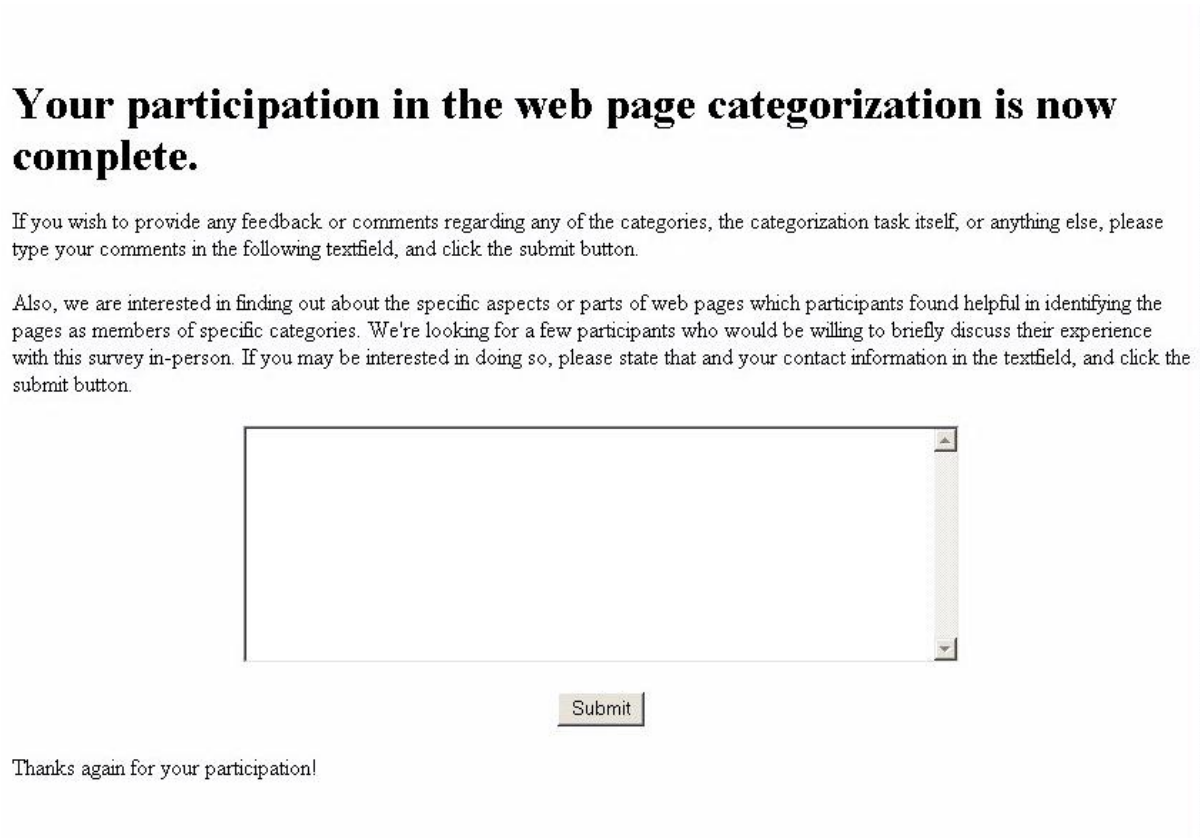
APPENDIX Q

SAMPLE SCREEN – GENRE PALETTE VALIDATION STUDY



APPENDIX R

FEEDBACK COLLECTION PAGE – GENRE PALETTE VALIDATION STUDY



Your participation in the web page categorization is now complete.

If you wish to provide any feedback or comments regarding any of the categories, the categorization task itself, or anything else, please type your comments in the following textfield, and click the submit button.

Also, we are interested in finding out about the specific aspects or parts of web pages which participants found helpful in identifying the pages as members of specific categories. We're looking for a few participants who would be willing to briefly discuss their experience with this survey in-person. If you may be interested in doing so, please state that and your contact information in the textfield, and click the submit button.

Submit

Thanks again for your participation!

APPENDIX S

PAGE 31 – GENRE PALETTE VALIDATION STUDY

Home » Admissions » Programs of Study & Courses » Columbia Publishing Course »

Application Information
- » Who Should Apply?
- » Frequently asked Questions
- » Fees
- » Application Forms

Columbia University
Graduate School of Journalism

The Columbia Publishing Course

Who Should Apply?

Search the Site:

directory | printer-friendly | small text | large text

The course is primarily aimed at recent college graduates. Other applicants are not discouraged, however. Many students have worked in publishing briefly and would like to broaden their understanding of the field, or have decided to make a career change from an unrelated field.

Students from the 2003 Publishing Course.

The choice of college major is incidental to acceptance, but applicants must have successfully completed all requirements for a B.A. or B.S. degree by June, 2004. Publishing is by no means restricted to the editorial function. While most applicants have majored in English and the humanities, many have majored in other disciplines.

Students with a demonstrated interest in publishing have always gained most directly from the course. Those who have worked on high school or college publications are familiar with publishing's long hours and constant deadline pressure. Likewise, those with bookstore, library, or office experience have skills and insights that publishers find valuable. Many types of interests, work, or volunteer experiences can be considered related to publishing; photography, graphic arts, sales, or marketing experience can qualify as valuable training for publishing.

The Columbia Publishing Course should not, however, be confused with a journalism program. Because of other educational opportunities available at Columbia and elsewhere, the course does not emphasize instruction in journalism or creative writing. Applicants with writing experience who seek new ways to apply their skills within the world of publishing--as editors, publicists, designers or publishers--are encouraged to apply.

Applicants should note that the Columbia Publishing Course is a highly intensive six-week program, with students expected to attend classes and workshops every morning, afternoon, and evening, as well as most weekends. With homework added to group sessions, students can expect little free time during the course.

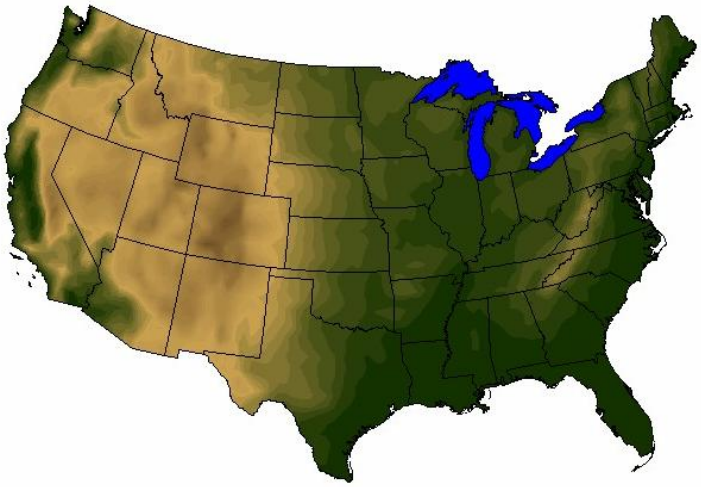
APPENDIX T

PAGE 69 – GENRE PALETTE VALIDATION STUDY

Weather By State

On this page, you can gain access to agricultural, climatic, observational, hydrometeorological, forecast, and model guidance data for all 50 U.S. states. Click on a state on the map below, or choose from the table of links below the map

NOTE: Point and click map works fine with Netscape - not so well with Internet Explorer. If using IE, use the text list below.



State		
Alabama	Louisiana	Ohio
Alaska	Maine	Oklahoma
Arizona	Maryland	Oregon
Arkansas	Massachusetts	Pennsylvania
California	Michigan	Rhode Island
Colorado	Minnesota	South Carolina
Connecticut	Mississippi	South Dakota
Delaware	Missouri	Tennessee
Florida	Montana	Texas
Georgia	Nebraska	Utah
Hawaii	Nevada	Vermont
Idaho	New Jersey	Virginia
Illinois	New Mexico	Washington
Indiana	New York	Washington DC
Iowa	New Hampshire	West Virginia
Kansas	North Carolina	Wisconsin
Kentucky	North Dakota	Wyoming

- Public Information Statements
- Last 5 Days of Record Weather Events

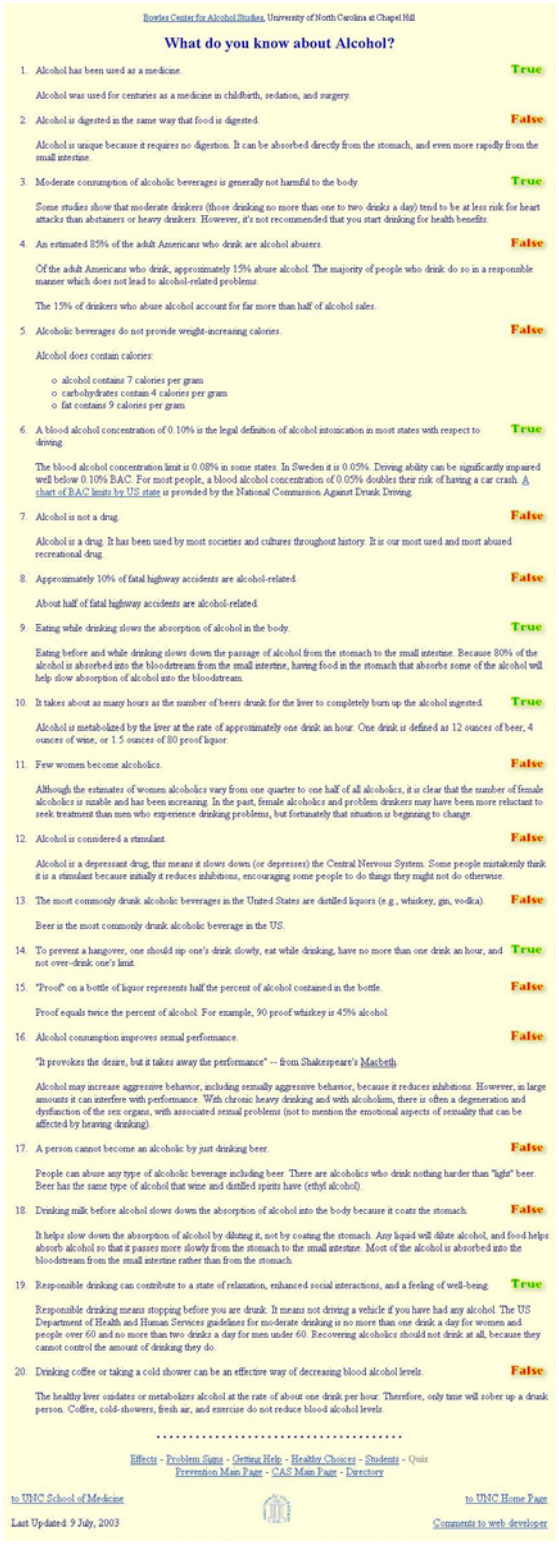
osuwx at geography.ohio-state.edu
The Ohio State University Atmospheric Sciences Program
Last updated on June 22, 2000 by Tony Schroeder

APPENDIX U

PAGE 122 – GENRE PALETTE VALIDATION STUDY

APPENDIX V

PAGE 154 – GENRE PALETTE VALIDATION STUDY



Bowles Center for Alcohol Studies, University of North Carolina at Chapel Hill

What do you know about Alcohol?

1. Alcohol has been used as a medicine. **True**

   Alcohol was used for centuries as a medicine in childbirth, sedation, and surgery.

2. Alcohol is digested in the same way that food is digested. **False**

   Alcohol is unique because it requires no digestion. It can be absorbed directly from the stomach, and even more rapidly from the small intestine.

3. Moderate consumption of alcoholic beverages is generally not harmful to the body. **True**

   Some studies show that moderate drinkers (those drinking no more than one to two drinks a day) tend to be at less risk for heart attacks than abstainers or heavy drinkers. However, it's not recommended that you start drinking for health benefits.

4. An estimated 85% of the adult Americans who drink are alcohol abusers. **False**

   Of the adult Americans who drink, approximately 15% abuse alcohol. The majority of people who drink do so in a responsible manner which does not lead to alcohol-related problems.

   The 15% of drinkers who abuse alcohol account for far more than half of alcohol sales.

5. Alcoholic beverages do not provide weight-increasing calories. **False**

   Alcohol does contain calories:

   - alcohol contains 7 calories per gram
   - carbohydrates contain 4 calories per gram
   - fat contains 9 calories per gram

6. A blood alcohol concentration of 0.10% is the legal definition of alcohol intoxication in most states with respect to driving. **True**

   The blood alcohol concentration limit is 0.08% in some states. In Sweden it is 0.05%. Driving ability can be significantly impaired well below 0.10% BAC. For most people, a blood alcohol concentration of 0.05% doubles their risk of having a car crash. A chart of BAC limits by US state is provided by the National Commission Against Drunk Driving.

7. Alcohol is not a drug. **False**

   Alcohol is a drug. It has been used by most societies and cultures throughout history. It is our most used and most abused recreational drug.

8. Approximately 10% of fatal highway accidents are alcohol-related. **False**

   About half of fatal highway accidents are alcohol-related.

9. Eating while drinking slows the absorption of alcohol in the body. **True**

   Eating before and while drinking slows down the passage of alcohol from the stomach to the small intestine. Because 80% of the alcohol is absorbed into the bloodstream from the small intestine, having food in the stomach that absorbs some of the alcohol will help slow absorption of alcohol into the bloodstream.

10. It takes about as many hours as the number of beers drunk for the liver to completely burn up the alcohol ingested. **True**

    Alcohol is metabolized by the liver at the rate of approximately one drink an hour. One drink is defined as 12 ounces of beer, 4 ounces of wine, or 1.5 ounces of 80 proof liquor.

11. Few women become alcoholics. **False**

    Although the estimates of women alcoholics vary from one quarter to one half of all alcoholics, it is clear that the number of female alcoholics is sizable and has been increasing. In the past, female alcoholics and problem drinkers may have been more reluctant to seek treatment than men who experience drinking problems, but fortunately that situation is beginning to change.

12. Alcohol is considered a stimulant. **False**

    Alcohol is a depressant drug, this means it slows down (or depresses) the Central Nervous System. Some people mistakenly think it is a stimulant because initially it reduces inhibitions, encouraging some people to do things they might not do otherwise.

13. The most commonly drunk alcoholic beverages in the United States are distilled liquors (e.g., whiskey, gin, vodka). **False**

    Beer is the most commonly drunk alcoholic beverage in the US.

14. To prevent a hangover, one should sip one's drink slowly, eat while drinking, have no more than one drink an hour, and not over-drink one's limit. **True**

15. "Proof" on a bottle of liquor represents half the percent of alcohol contained in the bottle. **False**

    Proof equals twice the percent of alcohol. For example, 90 proof whiskey is 45% alcohol.

16. Alcohol consumption improves sexual performance. **False**

    "It provokes the desire, but it takes away the performance" -- from Shakespeare's Macbeth.

    Alcohol may increase aggressive behavior, including sexually aggressive behavior, because it reduces inhibitions. However, in large amounts it can interfere with performance. With chronic heavy drinking and with alcoholism, there is often a degeneration and dysfunction of the sex organs, with associated sexual problems (not to mention the emotional aspects of sexuality that can be affected by heavy drinking).

17. A person cannot become an alcoholic by just drinking beer. **False**

    People can abuse any type of alcoholic beverage including beer. There are alcoholics who drink nothing harder than "light" beer. Beer has the same type of alcohol that wine and distilled spirits have (ethyl alcohol).

18. Drinking milk before alcohol slows down the absorption of alcohol into the body because it coats the stomach. **False**

    It helps slow down the absorption of alcohol by diluting it, not by coating the stomach. Any liquid will dilute alcohol, and food helps absorb alcohol so that it passes more slowly from the stomach to the small intestine. Most of the alcohol is absorbed into the bloodstream from the small intestine rather than from the stomach.

19. Responsible drinking can contribute to a state of relaxation, enhanced social interactions, and a feeling of well-being. **True**

    Responsible drinking means stopping before you are drunk. It means not driving a vehicle if you have had any alcohol. The US Department of Health and Human Services guidelines for moderate drinking is no more than one drink a day for women and people over 60 and no more than two drinks a day for men under 60. Recovering alcoholics should not drink at all, because they cannot control the amount of drinking they do.

20. Drinking coffee or taking a cold shower can be an effective way of decreasing blood alcohol levels. **False**

    The healthy liver oxidates or metabolizes alcohol at the rate of about one drink per hour. Therefore, only time will sober up a drunk person. Coffee, cold-showers, fresh air, and exercise do not reduce blood alcohol levels.

Effects - Problem Signs - Getting Help - Healthy Choices - Students - Quiz
Prevention Main Page - CAS Main Page - Directory

to UNC School of Medicine — to UNC Home Page

Last Updated: 9 July, 2003 — Comments to web developer

APPENDIX W

PAGE 158 – GENRE PALETTE VALIDATION STUDY

Understanding The 4-Year Plan
Researching Majors and Careers
Creating Career Connections
Improving Employability
Gaining Practical Experience
Attending Career Fairs
Applying to Graduate & Professional School
Frequently Asked Questions
Calendar of Events

SITE MAP

Search
Go
Google powered by

Career Timeline (4 Year Plan)

Freshman Year

- Seek career counseling.
- Explore various career fields.
- Get to know faculty and staff.
- Begin to participate in campus activities.
- Investigate experiential education.

Sophomore Year

- Declare a major.
- Seek leadership positions in various activities.
- Attend career and job fairs to get an idea of what employers are seeking.
- Consider enrolling in 2 credit hour Career Awareness Course (CAEN102)

Junior Year

- Register with the Career Center.
- Begin to conduct informational interviews with employers. Obtain career-related experience.
- Research graduate or professional schools and take entrance exams.
- Write a resume and participate in mock interviews conducted by the Career Center.

Senior Year

- Attend resume/interview workshops at the Career Center.
- Identify skills you can offer employers and further develop your resume.
- Register with the Career Center and complement this with your individual efforts.
- Attend job fairs and make employer contacts.
- Use the Career Center Resource Library to research job search strategies and potential employers.
- Send resume and cover letters and follow-up with telephone calls.

Please feel free to contact the office at:
(979)845-5139
Or come by John J. Koldus Building, Suite 209
and we'll be happy to help you!

TAMU Career Center · 1233 TAMU · College Station, TX 77843-1233
(979)845-5139 · (979)845-2979 (Fax)

Please send comments and suggestions regarding this site to: Webmaster@cctr.tamu.edu.

APPENDIX X

PAGE 208 – GENRE PALETTE VALIDATION STUDY

She Was Such A Great Person

My mother died yesterday (March 1, 2001) at the age of 46 from her second aneurysm in 5 years. In fall of 1996 she had her first aneurysm and was not supposed to live...she was out of the hospital a week after the surgery and was 100% normal again. It was if nothing had ever happened. Well, Wednesday morning she was getting ready for work and my father found her upstairs in the bathroom. 911 was called and she was taken to the hospital. She spent a day in ICU and after numerous tests on her brain, it was determined that she was 100% brain dead.

I don't even know what to say or do. It's unbelievable. She beat this same problem less than 5 years ago...now she's gone. It makes me want to throw up. She was such a great person. I love my mom so much...these next few days will be so painful, I just can't believe it. The blood wouldn't stop flowing in her brain and it just created so much pressure...they tried draining it...but she never regained consciousness. The doctor said that they almost didn't admit her to the hospital...which means she must have died on the way over and they brought her back to life.

In my own selfish ways, I'm feeling sorry for myself. But I know that my mother is in Heaven with God right now looking down on me. I last spoke with her Tuesday night, everything with her was fine. We talked about school and she told me she was proud of me. At the end of the conversation I told her I loved her and she told me she loved me too. That's it...that's the best final conversation I could ever have had with her. I'm so glad I told her that I loved her.

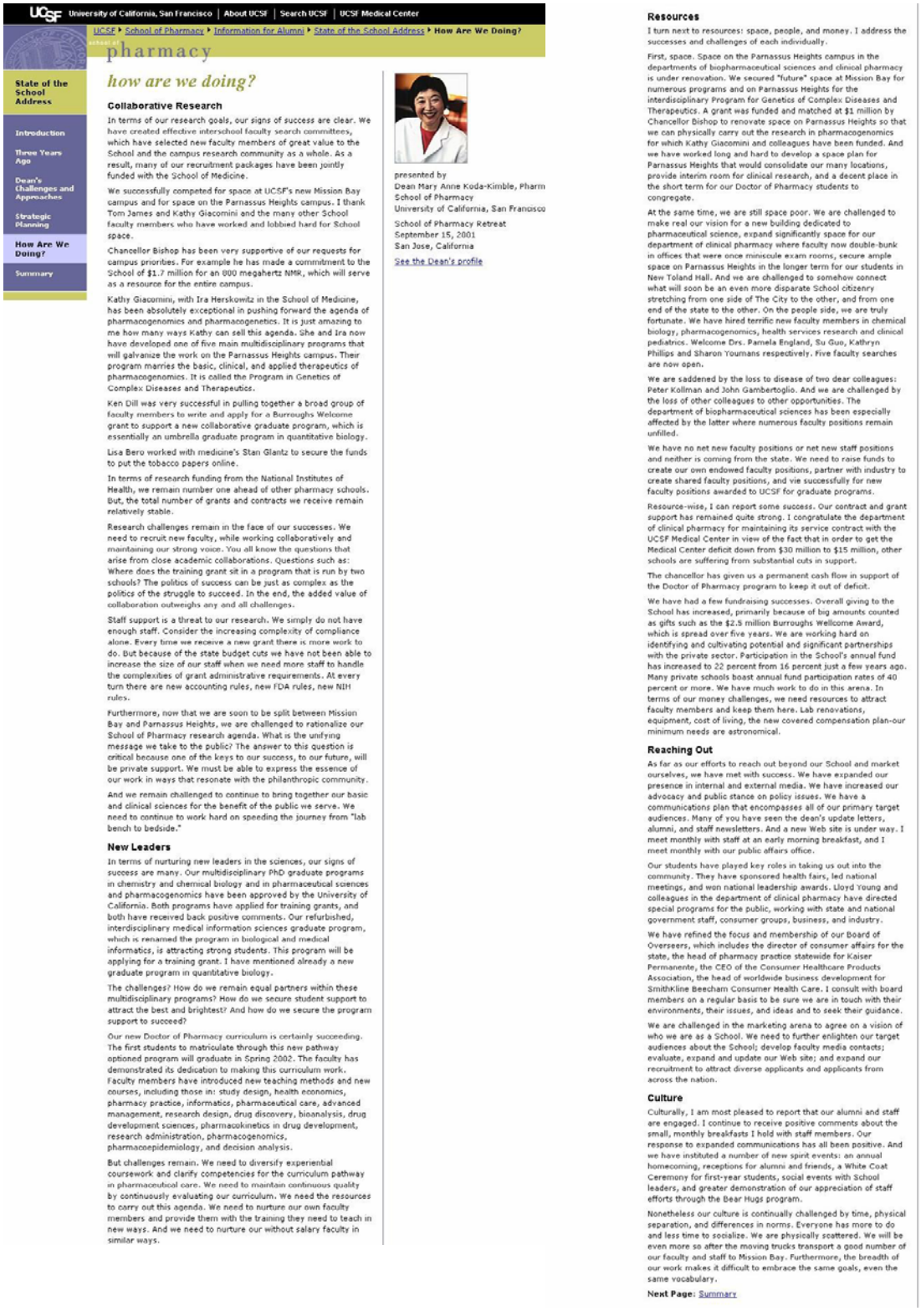
Discussion, comments, or questions: ***Nicholas Ploeger***

Return to contents

Return to Aneurysm & AVM Support

APPENDIX Y

PAGE 219 – GENRE PALETTE VALIDATION STUDY



UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center

UCSF ▸ School of Pharmacy ▸ Information for Alumni ▸ State of the School Address ▸ How Are We Doing?

school of pharmacy

State of the School Address

Introduction
Three Years Ago
Dean's Challenges and Approaches
Strategic Planning
How Are We Doing?
Summary

how are we doing?

Collaborative Research

In terms of our research goals, our signs of success are clear. We have created effective interschool faculty search committees, which have selected new faculty members of great value to the School and the campus research community as a whole. As a result, many of our recruitment packages have been jointly funded with the School of Medicine.

We successfully competed for space at UCSF's new Mission Bay campus and for space on the Parnassus Heights campus. I thank Tom James and Kathy Giacomini and the many other School faculty members who have worked and lobbied hard for School space.

Chancellor Bishop has been very supportive of our requests for campus priorities. For example he has made a commitment to the School of $1.7 million for an 800 megahertz NMR, which will serve as a resource for the entire campus.

Kathy Giacomini, with Ira Herskowitz in the School of Medicine, has been absolutely exceptional in pushing forward the agenda of pharmacogenomics and pharmacogenetics. It is just amazing to me how many ways Kathy can sell this agenda. She and Ira now have developed one of five main multidisciplinary programs that will galvanize the work on the Parnassus Heights campus. Their program marries the basic, clinical, and applied therapeutics of pharmacogenomics. It is called the Program in Genetics of Complex Diseases and Therapeutics.

Ken Dill was very successful in pulling together a broad group of faculty members to write and apply for a Burroughs Welcome grant to support a new collaborative graduate program, which is essentially an umbrella graduate program in quantitative biology.

Lisa Bero worked with medicine's Stan Glantz to secure the funds to put the tobacco papers online.

In terms of research funding from the National Institutes of Health, we remain number one ahead of other pharmacy schools. But, the total number of grants and contracts we receive remain relatively stable.

Research challenges remain in the face of our successes. We need to recruit new faculty, while working collaboratively and maintaining our strong voice. You all know the questions that arise from close academic collaborations. Questions such as: Where does the training grant sit in a program that is run by two schools? The politics of success can be just as complex as the politics of the struggle to succeed. In the end, the added value of collaboration outweighs any and all challenges.

Staff support is a threat to our research. We simply do not have enough staff. Consider the increasing complexity of compliance alone. Every time we receive a new grant there is more work to do. But because of the state budget cuts we have not been able to increase the size of our staff when we need more staff to handle the complexities of grant administrative requirements. At every turn there are new accounting rules, new FDA rules, new NIH rules.

Furthermore, now that we are soon to be split between Mission Bay and Parnassus Heights, we are challenged to rationalize our School of Pharmacy research agenda. What is the unifying message we take to the public? The answer to this question is critical because one of the keys to our success, to our future, will be private support. We must be able to express the essence of our work in ways that resonate with the philanthropic community.

And we remain challenged to continue to bring together our basic and clinical sciences for the benefit of the public we serve. We need to continue to work hard on speeding the journey from "lab bench to bedside."

New Leaders

In terms of nurturing new leaders in the sciences, our signs of success are many. Our multidisciplinary PhD graduate programs in chemistry and chemical biology and in pharmaceutical sciences and pharmacogenomics have been approved by the University of California. Both programs have applied for training grants, and both have received back positive comments. Our refurbished, interdisciplinary medical information sciences graduate program, which is renamed the program in biological and medical informatics, is attracting strong students. This program will be applying for a training grant. I have mentioned already a new graduate program in quantitative biology.

The challenges? How do we remain equal partners within these multidisciplinary programs? How do we secure student support to attract the best and brightest? And how do we secure the program support to succeed?

Our new Doctor of Pharmacy curriculum is certainly succeeding. The first students to matriculate through this new pathway optioned program will graduate in Spring 2002. The faculty has demonstrated its dedication to making this curriculum work. Faculty members have introduced new teaching methods and new courses, including those in: study design, health economics, pharmacy practice, informatics, pharmaceutical care, advanced management, research design, drug discovery, bioanalysis, drug development sciences, pharmacokinetics in drug development, research administration, pharmacogenomics, pharmacoepidemiology, and decision analysis.

But challenges remain. We need to diversify experiential coursework and clarify competencies for the curriculum pathway in pharmaceutical care. We need to maintain continuous quality by continuously evaluating our curriculum. We need the resources to carry out this agenda. We need to nurture our own faculty members and provide them with the training they need to teach in new ways. And we need to nurture our without salary faculty in similar ways.

presented by
Dean Mary Anne Koda-Kimble, Pharm
School of Pharmacy
University of California, San Francisco

School of Pharmacy Retreat
September 15, 2001
San Jose, California

See the Dean's profile

Resources

I turn next to resources: space, people, and money. I address the successes and challenges of each individually.

First, space. Space on the Parnassus Heights campus in the departments of biopharmaceutical sciences and clinical pharmacy is under renovation. We secured "future" space at Mission Bay for numerous programs and on Parnassus Heights for the interdisciplinary Program for Genetics of Complex Diseases and Therapeutics. A grant was funded and matched at $1 million by Chancellor Bishop to renovate space on Parnassus Heights so that we can physically carry out the research in pharmacogenomics for which Kathy Giacomini and colleagues have been funded. And we have worked long and hard to develop a space plan for Parnassus Heights that would consolidate our many locations, provide interim room for clinical research, and a decent place in the short term for our Doctor of Pharmacy students to congregate.

At the same time, we are still space poor. We are challenged to make real our vision for a new building dedicated to pharmaceutical science, expand significantly space for our department of clinical pharmacy where faculty now double-bunk in offices that were once miniscule exam rooms, secure ample space on Parnassus Heights in the longer term for our students in New Toland Hall. And we are challenged to somehow connect what will soon be an even more disparate School citizenry stretching from one side of The City to the other, and from one end of the state to the other. On the people side, we are truly fortunate. We have hired terrific new faculty members in chemical biology, pharmacogenomics, health services research and clinical pediatrics. Welcome Drs. Pamela England, Su Guo, Kathryn Phillips and Sharon Youmans respectively. Five faculty searches are now open.

We are saddened by the loss to disease of two dear colleagues: Peter Kollman and John Gambertoglio. And we are challenged by the loss of other colleagues to other opportunities. The department of biopharmaceutical sciences has been especially affected by the latter where numerous faculty positions remain unfilled.

We have no net new faculty positions or net new staff positions and neither is coming from the state. We need to raise funds to create our own endowed faculty positions, partner with industry to create shared faculty positions, and vie successfully for new faculty positions awarded to UCSF for graduate programs.

Resource-wise, I can report some success. Our contract and grant support has remained quite strong. I congratulate the department of clinical pharmacy for maintaining its service contract with the UCSF Medical Center in view of the fact that in order to get the Medical Center deficit down from $30 million to $15 million, other schools are suffering from substantial cuts in support.

The chancellor has given us a permanent cash flow in support of the Doctor of Pharmacy program to keep it out of deficit.

We have had a few fundraising successes. Overall giving to the School has increased, primarily because of big amounts counted as gifts such as the $2.5 million Burroughs Wellcome Award, which is spread over five years. We are working hard on identifying and cultivating potential and significant partnerships with the private sector. Participation in the School's annual fund has increased to 22 percent from 16 percent just a few years ago. Many private schools boast annual fund participation rates of 40 percent or more. We have much work to do in this arena. In terms of our money challenges, we need resources to attract faculty members and keep them here. Lab renovations, equipment, cost of living, the new covered compensation plan-our minimum needs are astronomical.

Reaching Out

As far as our efforts to reach out beyond our School and market ourselves, we have met with success. We have expanded our presence in internal and external media. We have increased our advocacy and public stance on policy issues. We have a communications plan that encompasses all of our primary target audiences. Many of you have seen the dean's update letters, alumni, and staff newsletters. And a new Web site is under way. I meet monthly with staff at an early morning breakfast, and I meet monthly with our public affairs office.

Our students have played key roles in taking us out into the community. They have sponsored health fairs, led national meetings, and won national leadership awards. Lloyd Young and colleagues in the department of clinical pharmacy have directed special programs for the public, working with state and national government staff, consumer groups, business, and industry.

We have refined the focus and membership of our Board of Overseers, which includes the director of consumer affairs for the state, the head of pharmacy practice statewide for Kaiser Permanente, the CEO of the Consumer Healthcare Products Association, the head of worldwide business development for SmithKline Beecham Consumer Health Care. I consult with board members on a regular basis to be sure we are in touch with their environments, their issues, and ideas and to seek their guidance.

We are challenged in the marketing arena to agree on a vision of who we are as a School. We need to further enlighten our target audiences about the School; develop faculty media contacts; evaluate, expand and update our Web site; and expand our recruitment to attract diverse applicants and applicants from across the nation.

Culture

Culturally, I am most pleased to report that our alumni and staff are engaged. I continue to receive positive comments about the small, monthly breakfasts I hold with staff members. Our response to expanded communications has all been positive. And we have instituted a number of new spirit events: an annual homecoming, receptions for alumni and friends, a White Coat Ceremony for first-year students, social events with School leaders, and greater demonstration of our appreciation of staff efforts through the Bear Hugs program.

Nonetheless our culture is continually challenged by time, physical separation, and differences in norms. Everyone has more to do and less time to socialize. We are physically scattered. We will be even more so after the moving trucks transport a good number of our faculty and staff to Mission Bay. Furthermore, the breadth of our work makes it difficult to embrace the same goals, even the same vocabulary.

Next Page: Summary

APPENDIX Z

EMAIL SOLICITATION – SEARCH RESULTS EVALUATION STUDY

Subject: 48 Needed: Help Improve Web Search, Get $10 Cash!
Date: Mon, 01 Nov 2004 12:59:05 -0500
From: Mark Rosso <rossom@meredith.edu>
To: emp_all@meredith.edu

Colleagues,

As many of you know, I am finishing up my dissertation research this year. I have one last experiment to run. I need 48 of you to help me.

Participants will be compensated $10 for completing five tasks involving scanning web search results and web pages, and judging their usefulness relative to a specific search task. The session will last approximately from 60-90 minutes, and will be held on-campus.

Volunteers should be currently employed as faculty or staff, and be familiar with using web search engines for searching for information on the Internet.

With this study, I hope to help further the development of new techniques for searching the World Wide Web. The study is part of doctoral research being done at the School of Information and Library Science at the University of North Carolina at Chapel Hill.

To apply, or for further information, please email me. You may want to include days and times that are typically convenient for you. Night and weekend appointments may also be possible.

Thanks,
Mark

--
Mark Rosso

Assistant Professor, Department of Mathematics and Computer Science
Meredith College, 3800 Hillsborough Street, Raleigh, NC 27607
(919) 760-2376 (phone) | (919) 760-8141 (fax)

APPENDIX AA

OPENING SCREEN – SEARCH RESULTS EVALUATION STUDY

Welcome to this Web Search Engine Study

These instructions pertain to the five tasks that you will be asked to do.

For each task, you will be shown a screen with a short search scenario. You are asked to read the scenario and imagine yourself in that search situation. When you are ready, you can then click the begin button, and you will be shown series of 20 search results that you might get from a web search engine during your search for the information you need.

For each result, rate the usefulness of the search result, relative to the solution of your imagined search situation, by clicking the appropriate button: "not useful at all", "slightly useful". "somewhat useful", and "highly useful". Make the decision as you normally would during a web search. The search situation will be on the screen at all times for you to refer to, if desired.

After rating the 20 search results, you will be shown a series of 20 web pages. Rate the usefulness of the page, again, relative to the solution of your search situation. The choices are the same as before: "not useful at all", "slightly useful", "somewhat useful", and "highly useful". However, this time, you will be able to browse the links on the page to help you make your decision. The linked page will open in a separate window (even if the link just points to another spot on the same page). Again, make your decision as you normally would during a web search.

Your first task will be a "practice" task. You will also have the opportunity to stop and ask questions about the practice task when you finished. Four more tasks will follow. You will be given special instructions before beginning the last task.

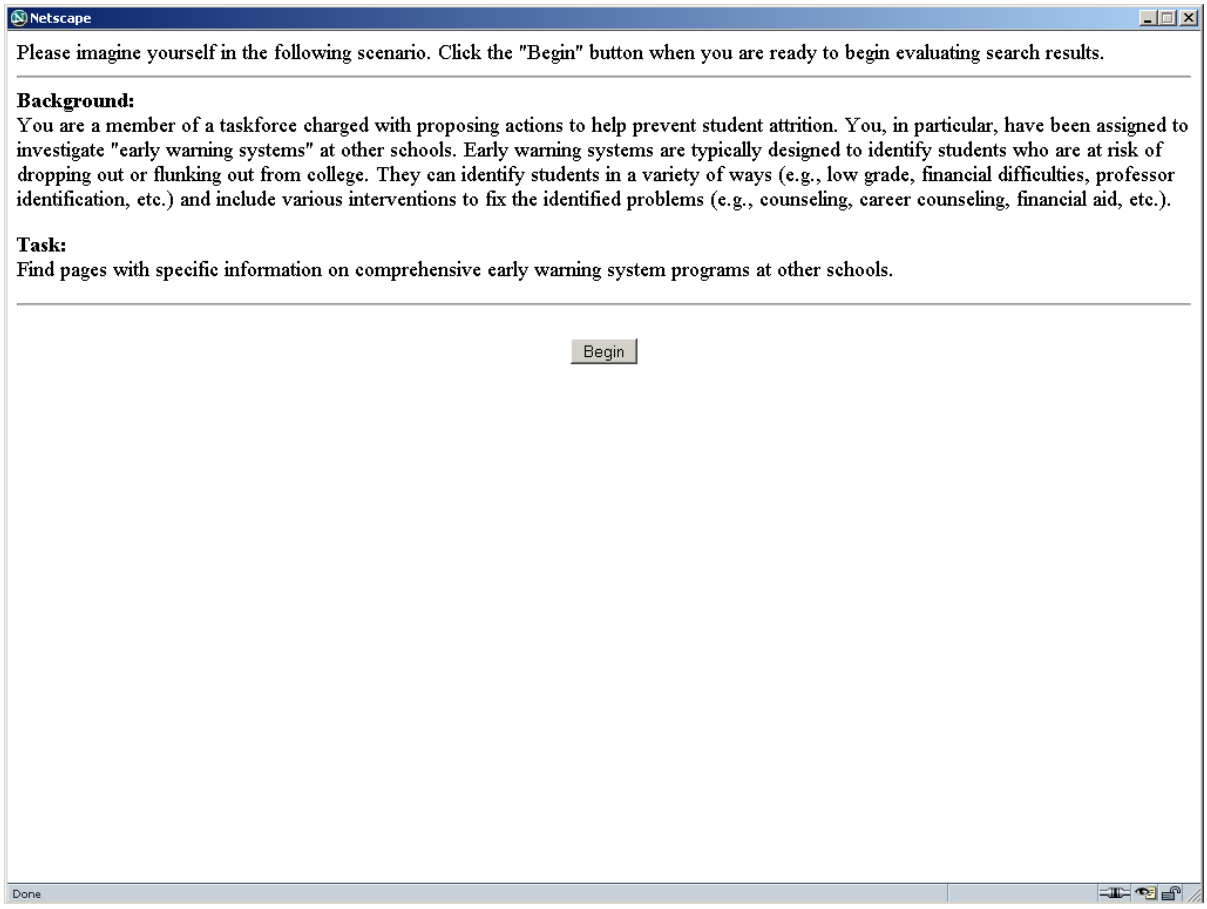
Any questions? If not, please click the first link when you are ready to begin the practice task.

Links To Tasks

- Practice Task
- Task 1
- Task 2
- Task 3
- Task 4

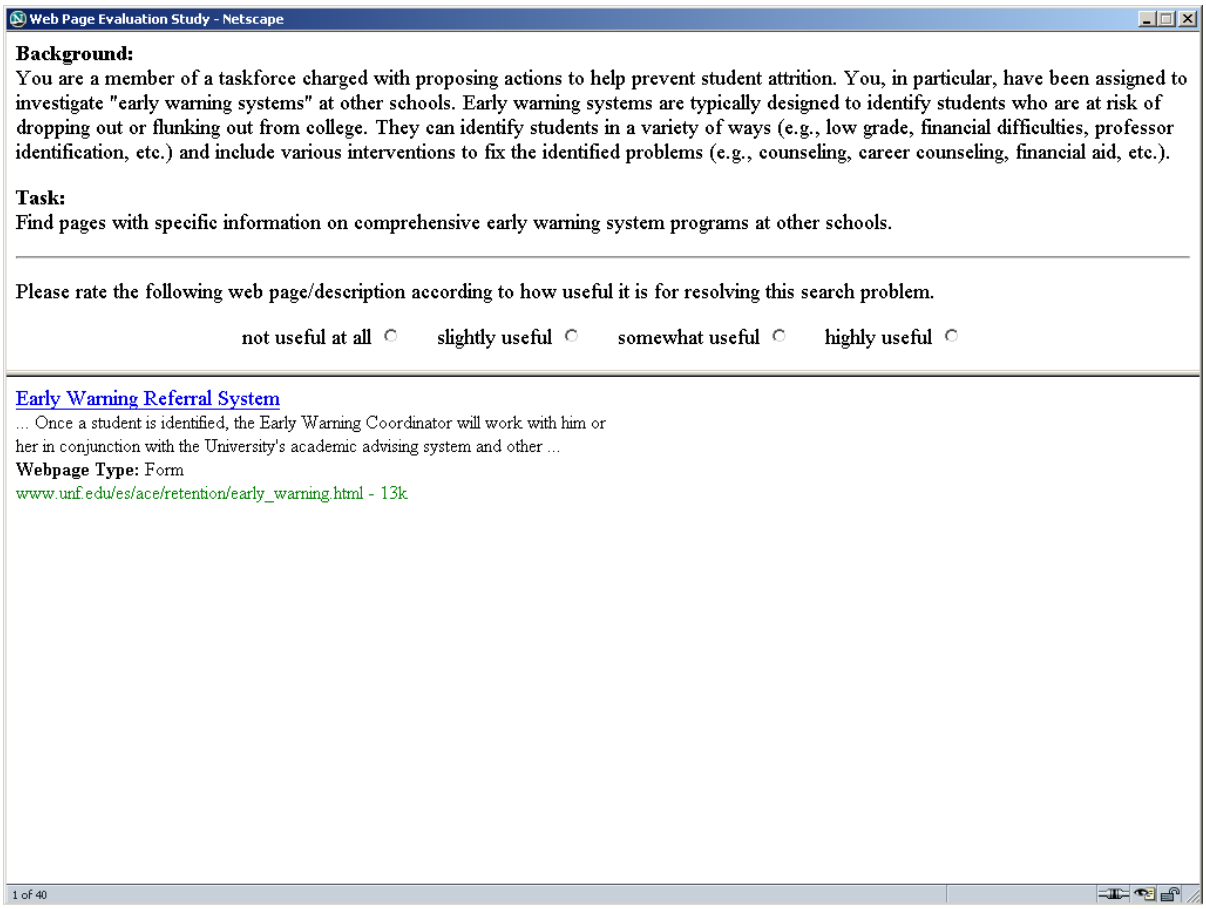
APPENDIX BB

TASK INFORMATION NEED SCREEN
SEARCH RESULTS EVALUATION STUDY



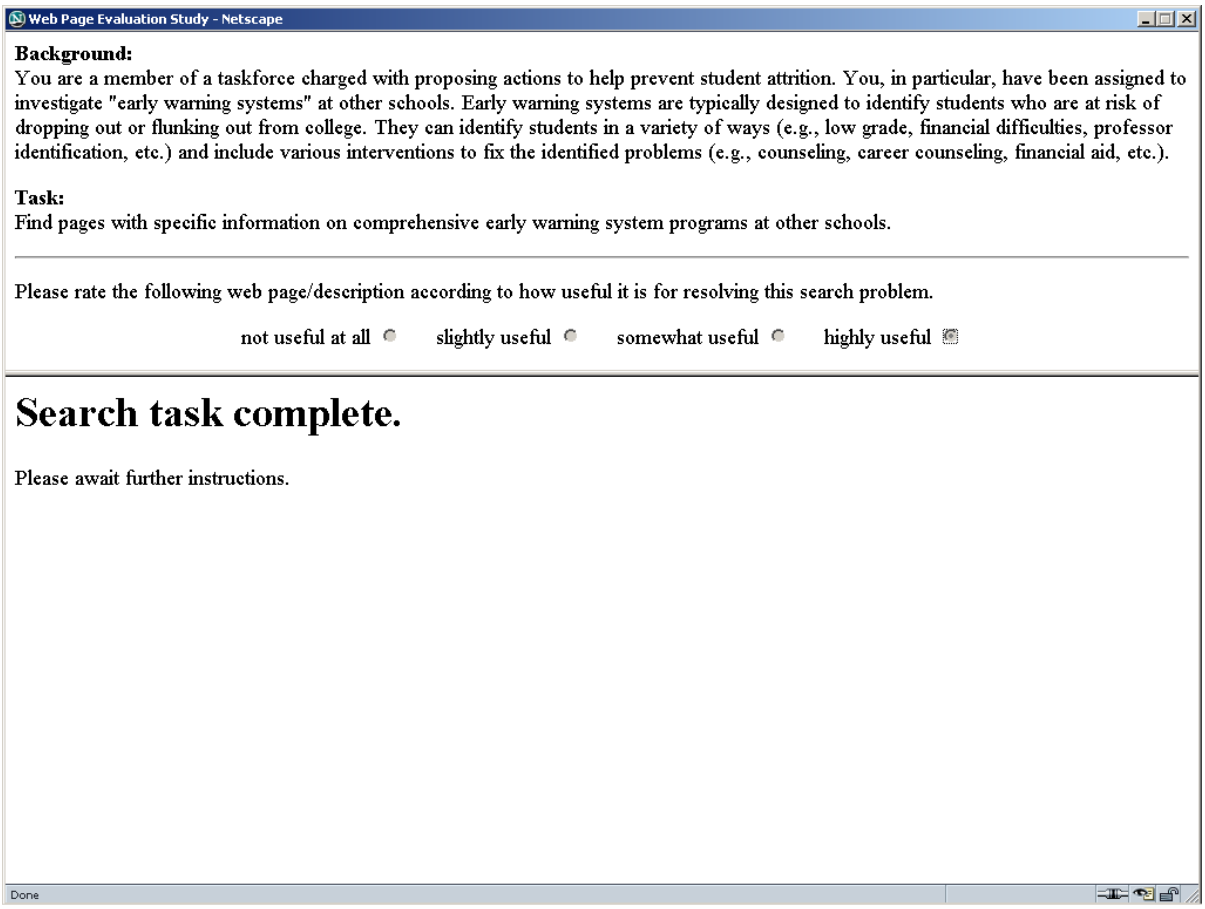
APPENDIX CC

SEARCH DESCRIPTION/WEBPAGE RATING SCREEN
SEARCH RESULTS EVALUATION STUDY



APPENDIX DD

TASK COMPLETE SCREEN
SEARCH RESULTS EVALUATION STUDY



APPENDIX EE

SEARCH RESULTS FOR EACH OF THE FIVE TASKS

Practice Task (no genres assigned to these pages)

Childcare Center Ground-Breaking
... Daycare groundbreaking-007 475 X 312 13 KB, Daycare groundbreaking-010 475 X 724 29 KB, Daycare groundbreaking-011 475 X 724 33 KB, Daycare groundbreaking-141 475 ...
www.uwrf.edu/buildingprojects/childslides/

Papers of the Barrett Daycare Center, 1935-
Guide to the Barrett Daycare Center Papers. [Special Collections] [Comments] [Credits] [Library Home Page] [URL: http://www.lib.virginia ...
www.lib.virginia.edu/speccol/exhibits/bdc/ - 4k

UCSD Daycare Oversight Committee
UCSD Daycare Oversight Committee Information. Table of Contents. ... Daycare choices at UCSD for employees and students. Infants through 5 1/2 years. ...
sam.ucsd.edu/daycare/ - 6k

UT Daycare Checklist Helps Ensure Child Safety
FOR IMMEDIATE USE April 15, 1999. UT Daycare Checklist Helps Ensure Child Safety (287). Parents should check day care centers for ...
pr.tennessee.edu/news/apr99/daycare.htm - 4k

Extended Daycare Center Breaks Ground - Childcare at UCI
... GSHIP and the Graduate Student GSHIP Survey Results OGS Launches New Website Moving In on Grad Student Housing Expansion of Extended Daycare GUEST COMMENTARY ...
www.rgs.uci.edu/gradvoice/articles/daycare.htm - 17k

Daycare Information - University of Maryland
Daycare Information. Each county in Maryland and the District of Columbia provide individualized referral services to accredited ...
www.faculty.umd.edu/relocate/daycare.html - 6k

Famiily Daycare Network--Bulletin Board
Home: Bulletin Boards: Family Daycare Provider Network. ... Vista. Family Daycare Provider Network. If you are a caregiver interested in: ...
info.med.yale.edu/chldstdy/parentsfirst/ bulletin_boards/bbs-familydaycarenetwork.html - 18k

NCR395 Liability Insurance and the Family Daycare Provider ...
... Liability Insurance and the Family Daycare Provider. ... Family daycare providers undertake great responsibility, and like most businesses, they need protection. ...
muextension.missouri.edu/xplor/regpubs/ncr395.htm - 9k

Daycare at Earlham College
Daycare at Earlham College. Earlham College, Richmond, Indiana 47374-4095. Earlham supports three daycare programs for children: The ...
www.earlham.edu/~daycare/ - 5k

[UT Office of the Dean of Students]
... Daycare Resource List. I. CHILDCARE CENTERS NEAR THE UT CAMPUS. ... II. OTHER USEFUL LINKS FOR DAYCARE CENTERS AND INFORMATION ABOUT KNOXVILLE. ...
web.utk.edu/~adultssc/daycare.html - 21k

[The Daycare Center - Ohio University Lancaster]
Realizing that convenient, drop-in child care is a necessity for students who are parents, OUL started a daycare service many years ago. ...
www.lancaster.ohiou.edu/student_services/daycare.shtml - 11k

[Chipola Junior College - Daycare]
A Developmental Preschool Program. Owner: Judy Jeter -- Director: Kristie Jeter. (850) 526-1112. Lic. # 65. Chipola Junior College has ...
www.chipola.edu/services/daycare/daycare1.htm

[Raffle Benefits Daycare Scholarships - MIT News Office]
... Raffle Benefits Daycare Scholarships. December 13, 2000. The Technology Children's Center, Inc. (TCC) is sponsoring a Razor Scooter ...
web.mit.edu/newsoffice/2000/tcc-1213.html

[phorum - Graduate Student Workshop: Do Babies Matter? - Drop in ...]
... Drop in daycare? Author: Anna-Lise Santella (---.dsl.emhril.ameritech.net) Date: 09-10-04 14:34 I'm searching for reliable drop-in ...
www.uchicago.edu/forum/read.php?f=148&i=40&t=40

[AUC | Faculty Services | Daycare Centers]
American University in Cairo. Faculty Services. Child Care—Daycare Centers. ... AUC Daycare, Stepping Stones, Small World Preschool, Stepping Stones, British Int. ...
www.aucegypt.edu/facstaff/facservices/daycare.html

[TE Job Card]
... Home Daycare assistant, # 500073. Work Study: preferred, Hours: 10-20 hours per week (variable). ... Business Name: patti pritiken family daycare. ...
www.hrs.wsu.edu/te/ jobcard.asp?jobnumber=500073&emptype=offcampus - 13k

[Delphi Questions: Daycare, educational time, PacifiCare]
Delphi Questions: Daycare, educational time, PacifiCare. ... In the past, UCAR has considered and rejected the idea of a corporate sponsored daycare center. ...
www.ucar.edu/communications/staffnotes/0104/delphi.html - 12k

[Waynesboro Senior Center and Adult Daycare Center]
Waynesboro Senior Center and Waynesboro Adult Daycare Center. The Waynesboro Senior Center began construction of it's terrace and ...
www.hort.vt.edu/human/WSCgarden.html - 2k

[WVU Parkersburg Children's Room daycare services]
WVU at Parkersburg The Children's Room. About the Children's Room. The Children's Room at West Virginia University at Parkersburg ...
www.wvup.edu/studentservices/Childrensroom.htm - 11k

daycare
Daycare is Finally Coming to MU! Students Vote to Support Referendum. The number of faculty ... fee increase to help. fund the daycare center. ...
www.marshall.edu/yearbook/tim/daycare.html - 4k

Task 1 (including genre annotations)

Middlesex County College - Finding a Student's Early Warning and ...
Academic Advisors' Manual Middlesex County College. Finding a Student's Early Warning and Midterm Grades. Starting in the XINQ screen ...
Webpage Type: FAQ/Help
www.middlesexcc.edu/manual/control.cfm/ID/1741 - 28k

AIU ASC Early Warning Program
... Academic Success Center/Title V Early Warning Program Procedure. ... advisors, or other appropriate staff are encouraged to complete a Student Referral Form after ...
Webpage Type: Welcome/Home Page
asc.alliant.edu/earlywarning/ - 24k

Myers University - College Achievement Program - CAP Information ...
College Achievement Program: Confidential Student Assessment and Early Warning. Please select or provide response's or information ...
Webpage Type: Form
www.dnmyers.edu/capeval.htm - 13k

University College Introduction to IUPUI
... instructor. Instructors may choose to count a student absent who comes late or leaves early. As ... F. The Early Warning System. National ...
Webpage Type: Welcome/Home Page
www.indiana.edu/~bltindy/introduction/ucoll.html - 7k

Red Flags for Early Warning
... Red Flags' for Early Warning. ... One might think that a poor grade would be warning enough, but ... Refer them to student services (see my home page http://dept.english ...
Webpage Type: Article
www.upenn.edu/almanac/v43/n23/redflags.html - 2k

Early Warning Referral System
... Once a student is identified, the Early Warning Coordinator will work with him or her in conjunction with the University's academic advising system and other ...
Webpage Type: Form
www.unf.edu/es/ace/retention/early_warning.html - 13k

WIU Undergraduate Catalog 2001-2002 -- Student Academic Progress
... Western Illinois University provides an early warning (mid-term) notification procedure to ... Any student who is permitted to register for classes at Western ...
Webpage Types: FAQ/Help; Instructional Materials
www.wiu.edu/catalog/01/progress.shtml - 9k

Early Warning Grades
... bottom to select "nothing to report - all student doing fine." That will indicate to anyone who checks that you have already turned in your early warning grades ...
Webpage Type: FAQ/Help
www.class.uidaho.edu/english/ comp_inst/Banner/Early_Warning_Grades.htm - 19k

LRNASST-L archives -- July 2002 (#41)
... College of Baltimore County) are piloting an Early Warning & Intervention Program with some great results! We used the Noel-Levitz Student Satisfaction Survey ...
Webpage Type: Forum/Interactive Discussion Archive
www.lists.ufl.edu/cgi-bin/wa?A2=ind0207& L=lrnasst-l&F=&S=&P=4266 - 9k

Frequently Asked Questions - NAUTICAL
... Up, ArgoNet Early Warning, Top. ... Please keep in mind, however, that the student will be locked out of COMPASS, and will not be able to maintain his/her account. ...
Webpage Type: FAQ/Help
nautical.uwf.edu/unitapp/faq/list.cfm?ID=85 - 13k

EWS
The Early Warning System (EWS) is an electronic alert system for instructors to ... The student's academic advisor and support services will also be notified so ...
Webpage Types: FAQ/Help; Welcome/Home Page
www.rpi.edu/dept/advising/EWS.html - 17k

WASC Accreditation Material
... Early Warning System. Half-sheet forms will be sent to the student's instructors (until an e-mail system can be developed) with the following three inquiries: ...
Webpage Type: Welcome/Home Page
wasc.csusb.edu/repository/STARS.htm - 17k

UC's McMicken College: For Faculty &Staff
... web site. Visit the registrar's web site and use your class list to flag a student for early warning. Early Intervention Frequently ...
Webpage Type: FAQ/Help
asweb.artsci.uc.edu/CollegeMain/ faculty_staff/earlywarn/index.cfm - 13k

UAF Adademic Advising Center: Early Warning Program Reports
The Early Warning program found many roadblocks in the way of access to student information that would enable them to easily and accurately determine the ...
Webpage Types: FAQ/Help; Welcome/Home Page
www.uaf.edu/advising/ew_reports/ - 9k

Resources for Faculty
... The student is referred to the Freshman Experience Office by faculty referral. The Freshman Early Warning System assists students in accessing needed academic ...
Webpage Type: FAQ/Help
www.asu.edu/provost/success/faculty.html - 12k

Academic Early Warning System
... in respective study group (all of the Academic Early Warning System courses ... from the Office of Academic Support Services will contact each student advising him ...
Webpage Types: Course List; Welcome/Home Page
www.albany.edu/eop/oass-eop/academic_early_warning.html - 3k

Shippensburg Student information
... Early Warning Grades Posted: 09-30-2004 10:15 Early Warning Grades are due on Friday ... close at 4:00 pm To schedule go to https://info.ship.edu/student/app/login ...
Webpage Types: Index/Table of Contents/Links; Welcome/Home Page
www.ship.edu/espfor/students/ - 35k

Early Warning Monitor
... prototype. [1] Graduate Student in the UNLV Health Physics Department. ... Vegas. Early Warning Monitor Presentation*. Early Warning Monitor Poster*. ...
Webpage Type: Article
nstg.nevada.edu/earlywarningmonitor.html - 10k

Division of Student Affairs - Loyola University New Orleans
... "What Qualifies A Leadership Program As A Success", Association of College Unions Bulletin, July 1992; "An Early Warning System Aids Student Affairs in ...
Webpage Type: Personal Website
www.loyno.edu/studentaffairs/mcneilresume.html - 24k

Task 2 (including genre annotations)

Prof. Terry Myers Zawacki
... nonfiction writing, freshman and advanced composition, writing ethnography, and the teaching of composition. She serves on the General Education, Composition ...
Webpage Types: Course List; Personal Website
classweb.gmu.edu/tzawacki/ - 6k

Assessment Office - Freshman Seminar Requirements - General ...
... student earns a C- or better. Writing Committee Guidelines for Freshman Seminars with a "W" Designation. Return to General Education.
Webpage Types: Course Descripton; Instructional Materials
www.wm.edu/wmoa/freshman_seminar_requirement.htm - 10k

General Education Curriculum
... The General Education Core Curriculum that follows is for ... I 010 - English 1320 - College Writing II 011 ... Biology Chemistry 1310 - Introductory Chemistry for Non ...
Webpage Type: Course List
www.txstate.edu/ucollege/gened.html - 6k

2004-2005 Bethel University College of Arts & Sciences Catalog, St ...
... 1, Freshman, ... 3, Senior, College Writing, Self-Awareness and Group Interaction course ... Bethel's General Education curriculum is a unified and developmental sequence ...
Webpage Type: Course List
cas.bethel.edu/catalog/acadinfo/curric/genreq.html - 26k

Resource Conservation -- Land and People Option
... BIOL 121N, Introductory Ecology, 3. BIOL 122N, Introductory Ecology Lab, 1. ... SUMMER - Between Freshman and Sophomore Year. ... FOR 220W, Technical Writing**, 2. ...
Webpage Type: Course List
www.forestry.umt.edu/academics/ Undergrad/ResCon/landandpeople.htm - 25k

CAS Incoming Freshman Home Page
... of Graduate and Undergraduate Studies Incoming Freshman College Orientation ... Most general education courses are 3 hours. ... are each 3 hours; Prep Writing and Lab ...
Webpage Types: Course List; FAQ/Help
www.cas.usf.edu/FallFreshman/coursesciences.html - 29k

General Education Requirments in UMB's CAS
... one Intermediate Seminar (G200 course) in the new General Education Curriculum (see ... to satisfy two writing requirements: the freshman writing requirement and ...
Webpage Type: FAQ/Help
www.umb.edu/academics/cla/gened.html - 31k

Indiana University Bloomington School of Nursing General Education ...
... in Natural & Mathematical Sciences; COAS S115 Freshman Seminar in ... W200 Using Computers in Education; ENG W103 Introductory Creative Writing; ENG W131 or ...
Webpage Type: Course List
www.indiana.edu/~iubnurse/home/course/cluster.html - 34k

General Education Courses
... EN208 Business/Technical Writing (3), PS120 Meteorology (3), SPECIAL - GENERAL EDUCATION - OPTIONS-. ... Beginning French II (4), Freshman Interest Groups ...
Webpage Type: Course List
www.washburn.edu/services/acadadv/gened2.html - 25k

General Education Electives Course Description
... Relevant assignments will develop students' critical, analytical and writing skills. ... An introductory audience-oriented examination of the elements of theatre ...
Webpage Types: Course Description; Course List
www.cs.odu.edu/~advisor/program/generalelectives.html - 43k

George Fox University: Course Catalogs: Undergraduate Catalog ...
... Pass/No Pass WRIT 110 Freshman Composition 3 hours. A course concentrating on expository writing, with an introduction to basic research methods. ...
Webpage Types: Course Description; Course List
www.georgefox.edu/catalog/ undergrad/catcourses/writ.html - 28k

General Education Task Force Draft Minutes
... freshman seminars. * Provides direct instruction. ... One 9 unit "Writing/Expression" course. ... EPICS (Engineering Practices Introductory Course Sequence). ...
Webpage Types: Article; Diary, Weblog or Blog
www.admin.mtu.edu/admin/ vpinst/minutes/14genedminutes.htm - 16k

Temple Times: Faculty Senate continues debate on general education
... studies requirement; to permit introductory courses to a major to also qualify for the general education program; to restructure writing-intensive courses; and ...
Webpage Type: Article
www.temple.edu/temple_times/3-4-04/senate.html - 13k

General Education Matrix - Truman State University
... several minor changes to its general education offerings in ... introduction to critical reading, writing, and thinking ... Freshman Program, (1 hour assigned else- where ...
Webpage Type: Course List
www.truman.edu/pages/267.asp - 44k

General Education, Ethnic Studies and Global Perspective ...
... 326-102 326-112, Freshman English -- Reading and Related Writing or Freshman English -- Honors II, 3. ... I. General Education Electives, 0-6 Credits. ...
Webpage Type: Course List
www.uwstout.edu/ugbulletin/ugb_gened.html - 19k

SUNY Institute of Technology - Registrar
... See description of the Freshman General Education Core ... Technique and Style History of American Art Theater and Communication Creative Writing Music Appreciation ...
Webpage Type: Course List
www.sunyit.edu/administration/ offices/registrar/?select=general_education - 24k

Class Schedule | University Core (GE + Religion) Detail
... Writing (to be completed during freshman year ... 300/H (has prerequisite)—Philosophical Writing PlSc 200 ... Group 1: Chem 101 —Introductory General Chemistry Chem ...
Webpage Type: Course List
saas.byu.edu/classSchedule/fall/geDetail.aspx - 64k

General Education Courses Declared Inactive as of 12/1/03
... of General Education, without regard to Introductory, or Further ... 27 syllabi came to the General Education Coordinator after ... found broad use of writing and oral ...
Webpage Type: Article
webs.wichita.edu/senate/GErpt04.htm - 73k

Sample Freshman/Sophomore Schedules—Geography
... Freshman Year. ... to University Life (1 hour); ENG 110 -- Writing I (3 ... Global Issues: Geographic Perspective (3 hours); GRY 142 -- Introductory Physical Geography (4 ...
Webpage Type: Course List
geosciences.smsu.edu/Geography/Schedules.htm - 11k

Writing Accross the Curriculum
... must complete four writing intensive (WI) general education courses, or allowable substitutions, in addition to the two required courses in freshman composition ...
Webpage Type: Course Description
www.wright.edu/gened/gewac.html - 13k

Task 3 (including genre annotations)

Social Science Undergraduate Student Guide
... Emporia State has an outstanding School of Liberal Arts and Sciences that provides all the courses necessary for pre-law instruction. ...
Webpage Type: Welcome/Home Page
www.emporia.edu/socsci/prelaw.html - 4k

Homepage
Juan Carlos Huerta, Ph.D. Associate Professor of Political Science 6300 Ocean Drive 360 Center for Instruction Corpus Christi, TX 78412 voice ... Pre-Law Advisor. ...
Webpage Type: Personal Website
www.tamucc.edu/~jhuerta/ - 4k

IPFW Pre-Law Handbook
... Moreover, the pre-law library located in my office contains a large number of law ... somewhat from school to school, the Socratic method of instruction has been ...
Webpage Type: FAQ/Help
www.ipfw.edu/pols/prelaw.htm - 18k

Departments of Instruction
... Science Psychology Physical Therapy Radiologic Sciences Reading Instruction Secondary Education ... The finest pre-law education is considered by many to be a four ...
Webpage Type: Welcome/Home Page
www.southalabama.edu/bulletin/bulletin9798/departmt.htm - 27k

Allied Health Information
... Home-LSAT: 56 hours of instruction either Online at $245.00 ... University of Chicago's Prelaw Guide. ... Akron: The Department of Political Science Pre-Law Advice Home ...
Webpage Type: Index/Table of Contents/Links
www.indiana.edu/~udivhpp/law.html - 19k

University of Nebraska-Lincoln - Scarlet's Web - 8/28/03
... So Dawes attended the Pre-Law Institute with the goal of getting as much ... In addition to the classroom instruction and brief seminars, tours, practice tests ...
Webpage Type: Article
www.unl.edu/scarlet/v13n20/v13n20features.html - 17k

MTU Department of Social Sciences - Undergraduate Programs - Pre ...
... The Pre-Law curriculum provides students with a substantial block of free electives (16 credits ... For information about courses, see Undergraduate Instruction.
Webpage Type: Welcome/Home Page
www.social.mtu.edu/ugprelaw.htm - 10k

Pre-Law
... a proper climate for this process is the quality of undergraduate instruction. ... At Gannon University individuals expressing an interest in Pre-Law are initially ...
Webpage Types: Course Description; Course List
www.gannon.edu/catalog/2003_2004/html/pre-law.html - 13k

Pierce College Catalog 2003-2005, Program listing
... LEGAL 275 Alternative Dispute Resolution (5). Total Credits Required 92. *Meets related instruction requirements for professional/technical programs. Pre-Law. ...
Webpage Type: Course List
www.pierce.ctc.edu/programs/programlist/law.php3 - 25k

University of Rhode Island News Releases
... configurations can be used for collaborative learning, as well as for seminar instruction. State-of-the-art electronic equipment, a pre-law advising center ...
Webpage Type: Article
www.uri.edu/news/releases/html/98-1203-01.htm - 13k

Advice For Freshmen and Sophomores
... Supplemental Instruction offerings) as a means of building in study time and skill development time. Spend some time exploring materials on pre-law preparation ...
Webpage Type: FAQ/Help
www.unm.edu/~pre/law/freshmanandsophomores.htm - 6k

WHY PRE
... students to do a traditional major, rather than a contrived "Pre-Law" major. ... Each of these majors at Queens has dedicated faculty and excellent instruction. ...
Webpage Type: Welcome/Home Page
campus.queens.edu/cas/acad_programs/ pre_law/Pre-Law/why_pre.htm - 5k

IU Southeast | Brochures | PreLaw
... Your IUS Prelaw Advisor can give you preparation ... faculty; Small class sizes for more personal instruction; ... more information regarding the pre-law program at ...
Webpage Types: FAQ/Help; Welcome/Home Page
www.ius.edu/brochures/AcademicPrograms/ ?brochure=prelaw&SchoolCode=SSCI - 14k

The Pre-law Program at SUNY Brockport
... programs at SUNY Brockport will provide exactly that kind of instruction. ... What the prelaw program at SUNY Brockport does is introduce ... Campus Pre-Law advisor. ...
Webpage Type: Welcome/Home Page
www.brockport.edu/prelaw/ - 8k

Pre-Law Curriculum
... Junior and Senior pre-law majors also are eligible for scholarships which include ... Moreover, student evaluations of instruction consistently grant high marks to ...
Webpage Type: Welcome/Home Page
www.louisiana.edu/Academic/LiberalArts/POLS/prelaw.html - 10k

Pre Law at Texas State University-San Marcos
... Philosophy Courses of Special Interest to Pre Law Students: ... Curricular Matters. Most prelaw students have heard of the "case method" of instruction. ...
Webpage Types: Course List; FAQ/Help; Welcome/Home Page
www.polisci.txstate.edu/pre-law/Pre-Law.html - 81k

CUC - English
... departments for pre-law students, is a member of the Northeast Association of Pre-Law Advisors, Inc. Student Placement for Writing Instruction First-year ...
Webpage Type: Welcome/Home Page
www.cuc.edu/academic/departments/english/ - 21k

KWC Catalog - Political Science
... The major sub-fields of instruction are political theory, American government, comparative ... A minor in political science and a minor in pre-law also are ...
Webpage Types: Course Description; Course List; Welcome/Home Page
www.kwc.edu/academic/catalog/2000/catpolsc.htm - 12k

Tutors - Dunster House
... Tutors. Keli Ballinger Resident Tutor in Careers and Wellness. Martin Bell Resident Pre-Law Advisor. ... Scott Sambur Resident Pre-Law Advisor. ...
Webpage Type: Index/Table of Contents/Links
hcs.harvard.edu/~dunster/tutors.php - 20k

SPLS Events Archive
... Pacific PreLaw Conference November 14, 10:30am-2 ... application materials and advice to pre-law students. ... Kaplan Center Practice tests and instruction offered by ...
Webpage Type: Index/Table of Contents/Links
www.stanford.edu/group/SPLS/archive/eventsarchive.html - 11k

Task 4 (including genre annotations)

The Chronicle Online - Book: grade inflation exists at Duke
... Is grade inflation a crisis at Duke and at other selective colleges? Former Duke statistics professor Valen Johnson thinks so - and will reveal his evidence in ...
Webpage Type: Article
www.chronicle.duke.edu/vnews/display.v/ART/2003/04/04/3e8d816e80a8a - 33k

CRA Grade Inflation ewp-mac/0004060
... Using a unique grade inflation methodology on actual ratings and evaluation data for 1,407 ... Access statistics for this paper at LogEc which is a part of the ...
Webpage Type: Article
econwpa.wustl.edu/eprints/mac/papers/0004/0004060.abs - 6k

PEDABLOGUE: Grade Inflation
... These statistics were shocking enough to raise everyone's concerns about grade inflation and we have since been conducting surveys, hosting conversations in ...
Webpage Types: Diary, Weblog or Blog; Forum/Interactive Discussion Archive
blogs.setonhill.edu/MikeArnzen/000850.html - 32k

Palinurus> Readings> Featured Controversy> Grade Inflation
... as causes for the decline of education, complaints against Grade Inflation have an air ... But backed by questionable statistics (which usually refer to the elite ...
Webpage Types: Article; Index/Table of Contents/Links; Welcome/Home Page
complit.rutgers.edu/palinurus/ - 28k

On grade inflation ...
... Statistics Indicate CU Grade Inflation, Cornell, 2002; Grade Inflation: It's Time to Face the Facts, Chronicle of Higher Education, 2001: Highly recommended; ...
Webpage Type: Index/Table of Contents/Links
www.mae.ufl.edu/~vql/grade.inflation.html - 7k

Faculty comments on plus/minus grading system
... It is my opinion that grades should be lowered and that the grade inflation registered in the statistics we published under item 1 begin in the eighties (long ...
Webpage Type: Index/Table of Contents/Links
www.ua.edu/academic/facsen/reports/resp.html - 22k

NCRVE MDS-1203 - REFERENCES
... of Education, National Center for Education Statistics. Ziomek, RL, & Svec, JC (1995). High school grades and achievement: Evidence of grade inflation. ...
Webpage Type: Index/Table of Contents/Links
ncrve.berkeley.edu/abstracts/ MDS-1203/MDS-1203-REFEREN.html - 8k

Chun Seng Yip
... Grade Inflation, Higher Education, Adverse Selection, Free-Riding, Collective Reputation. JEL: D82, I21, J71. Nedstat Basic - Free web site statistics.
Webpage Type: Personal Website
www.econ.upenn.edu/~yipcs/economics.htm - 10k

Lesa Hoffman :: Teaching
... While some instructors might view this as grade inflation, it is exactly the opposite. ... to fulfill a requirement within the major to take statistics, or may ...
Webpage Type: Instructional Materials
www.personal.psu.edu/faculty/l/r/lrh15/teaching/ - 12k

R. SAYLOR BRECKENRIDGE
... 6: I was talking with some students yesterday about the nature of grade inflation . ... education; 2) this is a CNN story referencing Honors statistics from Harvard ...
Webpage Type: Diary, Weblog or Blog
www.wfu.edu/users/breckers/ - 25k

Correspondence
... Grade Inflation and Date of Faculty hire. Schoolcraft College (Livonia, MI) faculty have lately been bombarded by administration with reams of statistics ...
Webpage Type: Forum/Interactive Discussion Archive
www.bus.lsu.edu/accounting/ faculty/lcrumbley/corespon.htm - 24k

Publications - Mohammed I. Chowdhury
... "Is there a Grade Difference in a Business Statistics Class? ... "Grade Inflation Revisited: An Empirical Approach," 27th Annual Conference of the International ...
Webpage Types: Index/Table of Contents/Links; Personal Website
www.emporia.edu/business/profiles/pub/chowdhum.htm - 9k

Nat'l Academies Press, Modern Interdisciplinary University ...
... Joint Statistical Meetings in which the Journal of Statistics Education managing ... Dartmouth, like many other schools, has wonderful grade inflation; for example ...
Webpage Types: Article; Product for Sale/Shopping
books.nap.edu/books/0309050332/html/117.html - 65k

The Student Life
... more venerable professors agree with the statistics. “I think it’s quite clear that there’s been some very significant grade inflation,” said Professor ...
Webpage Type: Article
www.tsl.pomona.edu/archives/02/1018/news/02.html - 20k

SCCC 312A – ProSeminar in Statistics
... in Statistics. Lecture Homework #8. Due Date: 3/17/2004 (Wednesday). Part I: (An application of the binomial distribution on deciding “grade inflation”) ...
Webpage Type: Instructional Materials
www.stat.sc.edu/~pena/SCCC312A/ LectureHomework/LectHW08.htm - 8k

Undergrad grade inflation? Reports may be overstated, study shows
... Education’s National Center for Education Statistics, most undergraduates ... available, suggest that concerns about widespread grade inflation in postsecondary ...
Webpage Type: Article
www.homepages.indiana.edu/ 092002/text/gradeinflation.html - 6k

Faculty File
... Statistics show that one department gave virtually nothing but A's and ... of recommendations aimed at addressing the core issues involved with grade inflation. ...
Webpage Type: Article
www.colby.edu/colby.mag/issues/85n3/faculty1.html - 9k

The Emperor's Good Grades
... A-. To combat this trend, Valen Johnson, a professor of statistics, has proposed ... These are the usual arguments for grade inflation, but they do not hold water ...
Webpage Type: Article
www.digitas.harvard.edu/ ~salient/issues/97comm/grades.h**tml - 7k**

Occasional Papers
... Statistics Indicate CU Grade Inflation (Cornell Daily Sun, April 11, 2002); Grade Inflation on the Rise (Cornell Daily Sun, April 15, 2002); ...
Webpage Type: Index/Table of Contents/Links
ece.gmu.edu/~wsutton/inflation/AAASoccasional.htm - 29k

PhD Professional Seminar -- 11 February 1997 -- Charles Van Loan
... The statistics generated are most likely meaningless. ... As professors respond to these pressures they contribute to grade inflation. ...
Webpage Type: Instructional Materials
www.cs.cornell.edu/degreeprogs/ ugrad/uphilos/DegradingGrading/ - 13k

APPENDIX FF

ORAL CONSENT FORM – SEARCH RESULTS EVALUATION STUDY

Web Search Results Evaluation Study

Thank you for responding to our invitation to participate in our study of Web Search Results Evaluation. This research is intended to help further efforts to make searching for information on the World Wide Web easier and more effective.

The work is being conducted by Mark Rosso, a doctoral student in the School of Information and Library Science at the University of North Carolina at Chapel Hill. If you have any questions about this study, please call the faculty advisor for this research, Dr. Stephanie Haas at (919) 962-8360.

If you choose to participate, you will be asked to perform five tasks. In each task, you will be shown a search scenario, and then asked to evaluate a series of search results and web pages, in accordance with how useful they are for resolving the search problem. The study should take approximately 60-90 minutes, and must be completed at one time. Of course, your consent does not preclude you from withdrawing your participation at any time.

Note: Your participation is anonymous. Your name will not be recorded or associated with the data in any way. With your permission, the final task will be tape-recorded.

The Behavioral Institutional Review Board of the University of North Carolina at Chapel Hill has approved this study. If you have any concerns about your rights as a participant in this study, you may contact the Board at 919-962-7761 or at aa-irb@unc.edu.

APPENDIX GG

ENTRY QUESTIONNAIRE – SEARCH RESULTS EVALUATION STUDY

Entry Questionnaire Participant # _____

1. Is your current position primarily staff or faculty?

2. How many years (lifetime) have you spent as faculty?

3. How many years (lifetime) have you spent as staff?

4. How many total hours/week do you spend using the web?

5. How many hours/week do you spend searching the web?

6. How many hours/week do you spend searching the web as part of a job-related activity?

7. Have you ever searched for information on any of the following topics?

   a. provision of daycare by colleges and universities
   b. college composition courses
   c. “early warning systems” for preventing student attrition
   d. pre-law curricula
   e. grade inflation

   Please briefly describe the situation motivating the search: what exactly were you looking for, and what was the intended use for the information?

APPENDIX HH

RELEVANCE DECISIONS BY GENRE BY PRESENCE OF GENRE ANNOTATION

ARTICLE

W/ GENRE — Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	37	8	6	8	59	62.7%	0.75	0.75
2	34	11	14	6	65	16.9%	0.92	(0.12)
3	25	14	19	12	70	27.1%	1.09	(0.74)
4	19	8	3	16	46	34.8%	1.65	(1.65)
Total	115	41	42	42	240	34.6%	1.07	(0.38)

PLAIN — Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	36	10	4	4	54	66.7%	0.56	0.56
2	26	18	6	10	60	30.0%	0.87	0.00
3	31	20	8	13	72	11.1%	1.32	(0.96)
4	17	14	4	19	54	35.2%	1.54	(1.54)
Total	110	62	22	46	240	33.8%	1.08	(0.51)

COURSE DESCRIPTION

W/ GENRE — Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	3	0	0	1	4	75.0%	0.75	0.75
2	7	6	1	8	22	27.3%	1.09	0.45
3	4	10	4	12	30	13.3%	1.00	(0.20)
4	4	5	7	24	40	60.0%	0.73	(0.73)
Total	18	21	12	45	96	38.5%	0.90	(0.23)

PLAIN — Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	2	2	3	2	9	22.2%	1.56	1.56
2	4	4	3	9	20	20.0%	1.25	0.85
3	6	10	3	10	29	10.3%	1.10	(0.41)
4	1	4	10	22	37	59.5%	0.57	(0.57)
Total	13	20	19	43	95	32.6%	0.97	(0.02)

COURSE LIST

W/ GENRE — Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	28	20	7	9	64	43.8%	0.95	0.95
2	44	37	10	23	114	32.5%	0.88	0.11
3	24	25	10	21	80	12.5%	1.18	(0.65)
4	11	9	8	18	46	39.1%	1.28	(1.28)
Total	107	91	35	71	304	30.6%	1.03	(0.13)

PLAIN Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	28	13	10	19	70	40.0%	1.29	1.29
2	34	23	16	20	93	24.7%	0.97	0.24
3	30	27	12	21	90	13.3%	1.20	(0.73)
4	9	10	10	18	47	38.3%	1.21	(1.21)
Total	101	73	48	78	300	27.0%	1.15	(0.04)

DIARY, WEBLOG or BLOG

W/ GENRE Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	8	1	2	2	13	61.5%	0.85	0.85
2	2	2	5	4	13	15.4%	1.15	0.85
3	1	3	4	6	14	28.6%	0.79	0.07
4	5	1	0	2	8	25.0%	2.13	(2.13)
Total	16	7	11	14	48	33.3%	1.13	0.13

PLAIN Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	5	1	0	2	8	62.5%	0.88	0.88
2	1	7	7	1	16	43.8%	0.63	0.50
3	4	2	4	4	14	28.6%	1.00	(0.43)
4	1	4	4	1	10	10.0%	1.50	(1.50)
Total	11	14	15	8	48	35.4%	0.96	(0.13)

FAQ/HELP

W/ GENRE Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	34	13	5	7	59	57.6%	0.75	0.75
2	21	13	9	15	58	22.4%	1.03	0.31
3	22	14	9	21	66	13.6%	1.20	(0.56)
4	16	11	11	19	57	33.3%	1.42	(1.42)
Total	93	51	34	62	240	31.3%	1.10	(0.23)

PLAIN Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	31	10	0	8	49	63.3%	0.69	0.69
2	21	25	14	15	75	33.3%	0.87	0.31
3	18	15	17	26	76	22.4%	1.01	(0.33)
4	8	8	5	19	40	47.5%	1.13	(1.13)
Total	78	58	36	68	240	38.3%	0.92	(0.05)

FORM

W/ GENRE

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	0	2	1	0	3	0.0%	1.33	1.33
2	3	5	2	0	10	50.0%	0.50	(0.10)
3	0	3	3	2	8	37.5%	0.63	(0.13)
4	0	4	5	2	11	18.2%	1.18	(1.18)
Total	3	14	11	4	32	31.3%	0.84	(0.34)

PLAIN

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	1	1	3	1	6	16.7%	1.67	1.67
2	1	2	1	4	8	25.0%	1.25	1.00
3	1	1	2	3	7	28.6%	0.86	0.00
4	0	4	3	3	10	30.0%	1.10	(1.10)
Total	3	8	9	11	31	25.8%	1.19	0.23

FORUM/DISCUSSION ARCHIVE

W/ GENRE

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	6	1	0	0	7	85.7%	0.14	0.14
2	6	3	1	3	13	23.1%	1.00	0.08
3	4	4	4	3	15	26.7%	1.00	(0.60)
4	5	4	1	3	13	23.1%	1.85	(1.85)
Total	21	12	6	9	48	33.3%	1.10	(0.65)

PLAIN

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	2	1	0	0	3	66.7%	0.33	0.33
2	3	2	2	1	8	25.0%	0.88	0.13
3	9	4	4	5	22	18.2%	1.23	(0.77)
4	5	5	4	1	15	6.7%	1.93	(1.93)
Total	19	12	10	7	48	18.8%	1.33	(0.92)

INDEX/TOC/LINKS

W/ GENRE

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	39	9	2	2	52	75.0%	0.37	0.37
2	19	6	12	5	42	14.3%	0.98	0.07
3	10	8	6	12	36	16.7%	1.11	(0.44)
4	7	6	4	13	30	43.3%	1.23	(1.23)
Total	75	29	24	32	160	40.0%	0.86	(0.19)

PLAIN Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	36	5	6	5	52	69.2%	0.62	0.62
2	16	9	6	4	35	25.7%	0.86	(0.06)
3	15	5	8	10	38	21.1%	1.18	(0.66)
4	6	8	4	17	35	48.6%	1.09	(1.09)
Total	73	27	24	36	160	43.8%	0.91	(0.21)

OTHER INSTRUCTIONAL MATERIALS

W/ GENRE Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	27	2	1	0	30	90.0%	0.13	0.13
2	12	3	0	2	17	17.6%	0.94	(0.47)
3	10	7	3	1	21	14.3%	1.33	(1.24)
4	3	6	1	2	12	16.7%	1.83	(1.83)
Total	52	18	5	5	80	43.8%	0.88	(0.65)

PLAIN Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	21	2	3	0	26	80.8%	0.31	0.31
2	18	7	4	0	29	24.1%	0.76	(0.48)
3	9	3	1	0	13	7.7%	1.62	(1.62)
4	3	3	2	4	12	33.3%	1.42	(1.42)
Total	51	15	10	4	80	41.3%	0.85	(0.55)

PERSONAL WEBSITE

W/ GENRE Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	18	5	5	4	32	56.3%	0.84	0.84
2	7	5	5	5	22	22.7%	1.00	0.36
3	6	8	2	2	18	11.1%	1.22	(1.00)
4	3	4	0	1	8	12.5%	2.13	(2.13)
Total	34	22	12	12	80	32.5%	1.10	0.00

PLAIN Full-text Rating

Summary Rating	1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	9	3	4	12	28	32.1%	1.68	1.68
2	16	4	3	4	27	14.8%	1.00	(0.19)
3	6	6	3	0	15	20.0%	1.20	(1.20)
4	6	1	1	1	9	11.1%	2.33	(2.33)
Total	37	14	11	17	79	21.5%	1.43	0.04

PRODUCT FOR SALE/SHOPPING

W/ GENRE

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	3	0	0	0	3	100.0%	0.00	0.00
2	5	1	0	0	6	16.7%	0.83	(0.83)
3	3	2	0	0	5	0.0%	1.60	(1.60)
4	2	0	0	0	2	0.0%	3.00	(3.00)
Total	13	3	0	0	16	25.0%	1.19	(1.19)

PLAIN

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	4	1	0	0	5	80.0%	0.20	0.20
2	5	0	0	0	5	0.0%	1.00	(1.00)
3	4	0	0	1	5	0.0%	1.80	(1.40)
4	1	0	0	0	1	0.0%	3.00	(3.00)
Total	14	1	0	1	16	25.0%	1.13	(0.88)

WELCOME/HOMEPAGE

W/ GENRE

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	16	10	5	15	46	34.8%	1.41	1.41
2	17	12	12	30	71	16.9%	1.25	0.77
3	18	20	14	30	82	17.1%	1.05	(0.32)
4	14	11	13	50	88	56.8%	0.88	(0.88)
Total	65	53	44	125	287	32.1%	1.10	0.06

PLAIN

Summary Rating	Full-text Rating 1	2	3	4	Total	Indicativity	Absolute Chg	Relative Chg
1	13	12	5	12	42	31.0%	1.38	1.38
2	14	24	12	21	71	33.8%	0.96	0.56
3	13	15	34	34	96	35.4%	0.78	(0.07)
4	7	12	16	44	79	55.7%	0.77	(0.77)
Total	47	63	67	111	288	39.9%	0.91	0.10

BIBLIOGRAPHY

Amitay, E. (2001) *What lays in the layout*. Ph.D. dissertation. Macquarie University. Retrieved on February 12, 2005 from http://www.ics.mq.edu.au/~einat/thesis/

Amitay, E. & Paris, C. (2000). Automatically summarising web sites—is there a way around it? *Proceedings of 9th Internet Conference on Information and Knowledge Management (CIKM 2000)*. 173-179.

Askehave, I. & Swales, J. (2001). Genre, identification and communicative purpose: a problem and a possible solution. *Applied Linguistics*, 22(2), 195-212.

Barry, C. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science*, 45(3), 149–159.

Barry, C. (1998). Document representations and clues to document relevance. *Journal of the American Society for Information Science*, 49(14), 1293–1303.

Barry, C., & Schamber, L. (1998). User criteria for relevance evaluation: a cross-situational comparison. *Information Processing & Management*, 34(2/3), 219–236.

Barsalou, L. (1983). Ad hoc categories. *Memory & Cognition*, 11(3), 211-227.

Bateman, J. (1998). Changes in relevance criteria: A longitudinal study. *Proceedings of the 61st American Society for Information Science Annual Meeting*, 23–32.

Beghtol, C. (2000). The concept of genre and its characteristics. *Bulletin of the American Society for Information Science and Technology*, 27(2), 17-19.

Belkin, N. & Muresan, G. (2004). Measuring web search effectiveness: Rutgers at interactive TREC. *World Wide Web 2004 Conference Workshop*. Retrieved on December 30, 2004 from http://www.scils.rutgers.edu/~muresan/Publications/wwwWsBelkin2004.pdf

Berkenkotter, C. & Huckin, T. (1993). Rethinking genre from a sociocognitive perspective. *Written Communication*, 10(4), 475-509.

Berlin, B., Breedlove, D. & Raven, P. (1974). *Principles of Tzeltal Plant Classification.* New York: Academic Press.

Bhatia, V. (1993). *Analysing genre: Language use in professional settings.* London: Longman.

Biber, D. (1988). *Variations Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3-43.

Biber, D. (1994). An analytical framework for register studies. In: *Sociolinguistic Perspectives on Register* (D. Biber and E. Finnegan, Eds.), pp. 31-56, New York: Oxford University Press, 1994.

Björneborn, L. (2004). *Small-world link structures across an academic web space : a library and information science approach.* Ph.D. dissertation. Royal School of Library and Information Science. Retrieved on November 29, 2004 at: http://www.db.dk/lb/phd/phd-thesis.pdf

Blair, D. (2002a). The challenge of commercial document retrieval, Part I: major issues, and a framework based on search exhaustivity, determinacy of representation and document collection size. *Information Processing & Management,* 38(2), 273-291.

Blair, D. (2002b). The challenge of commercial document retrieval, Part II: a strategy for document searching based on identifiable document partitions. *Information Processing & Management,* 38(2), 293-304.

Borlund, P. (2003a). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research,* 8(3), paper no. 152 [Available at: http://informationr.net/ir/8-3/paper152.html]

Borlund, P. (2003b). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913-925.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71-90.

Buckley, C., Salton, G., Allan, J. & Singhal, A. (1995). Automatic query expansion using SMART: TREC 3. *Proceedings of Third Text Retrieval Conference (TREC-3), NIST Special Publication 500-225,* 69-80.

Chen, H., Schuffels, C., & Orwig R. (1996). Internet categorization and search: a self-organizing approach. *Journal of Visual Communication and Image Representation*, 7(1), 88-102.

Church, K. & Rau, L. (1995). Commercial applications of natural language processing. *Communications of the ACM,* 38(11), 71-79.

Cleverdon, C. (1997). The Cranfield tests on index language devices. In K. S. Jones & P. Willett Eds.), Readings in information retrieval (pp. 47–59). San Francisco: Morgan Kaufman, Reprinted from Aslib Proceedings, (19), 173–192.

Cool, C., Belkin, N., Kantor, P., & Frieder, O. (1993). Characteristics of texts affecting relevance judgments. *Proceedings of the 14th National Online Meeting,* 77–84.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing & Management*, 36, 533–550.

Crowston, K. & Kwasnik, B. (2004). A framework for creating a facetted classification for genres: addressing issues of multidimensionality. *Proceedings of the 38nd Annual Hawaii International Conference on Systems Sciences, Digital Media Track*, IEEE Computer Society Press.

Crowston, K. & Kwasnik, B. (2003). Can document-genre metadata improve information access to large digital collections?, *Library Trends*, 52(2), 345-361.

Crowston, K., & Williams, M. (2000). Reproduced and emergent genres of communication on the World-Wide-Web, *The Information Society*, 16(3), 201-216.

Crowston, K., & Williams, M. (1997). Reproduced and emergent genres of communication on the World-Wide-Web, *Proceedings of the 30th Annual Hawaii International Conference on System Sciences: Digital Documents*, 30-39.

Dewe, J., Bretan, I., & Karlgren, J. (1998). Assembling a balanced corpus from the Internet. *Proceedings of the 11th Nordic Computational Linguistics Conference*, 100-107.

Dillon, A. (2000). Spatial-semantics: How users derive shape from information space. *Journal of the American Society for Information Science*, 51(6), 521-528.

Dillon. A. (1991). Readers' models of text structures: The case of academic articles. *International Journal of Man-Machine Studies*, 35, 913-925.

Dillon, A. & Gushrowski, B. (2000). Genres and the Web: Is the personal home page the first uniquely digital genre?, *Journal of the American Society for Information Science*, 5, 202-205.

Dillon, A., & Schaap, D. (1996). Expertise and the perception of shape in information. *Journal of the American Society for Information Science*, 47, 786–788.

Dillon, A., & Vaughan, M. (1997). "It's the journey and the destination": Shape and the emergent property of genre in evaluating digital documents. *New Review of Hypermedia & Multimedia*, 3, 91–106.

Dublin Core Metadata Initiative. (n.d.). Retrieved on January 20, 2003 from http://dublincore.org/

Dublin Core Resource Types (1998). Retrieved on August 18, 2002 from: http://sunsite.berkeley.edu/Metadata/structuralist.html

Fayol, M. (1991). *Text typologies: a cognitive approach*. In G. Denhiere and J. Rossi (eds.) *Text and Text Processing*, pp. 61-76. New York: Elsevier Science Publishers.

Ferguson, C. (1994). Dialect, register and genre: Working assumptions about conventionalization. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 15-30). New York: Oxford University Press.

Fidel, R. (1991). Searcher's selection of search keys, part I. *Journal of the American Society for Information Science*, 42(7), 490-500.

Foskett, D.J. (1972). A note on the concept of "relevance." *Information Storage and Retrieval*, 8(2), 77–78.

Francis, W. & Kucera, H. (1982). *Frequency Analysis of English Usage*. Houghton Mifflin Co.: New York.

Freedman, A & Medway, P. (1994). Locating genre studies: antecedents and prospects. In A. Freedman & P. Medway (eds.), *Genre and the New Rhetoric*, pp. 1-10. London: Taylor & Francis.

Furnas, G., Landauer, T., Gomez, L. & Dumais, S. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11), 964-971.

Giltrow, J. (1994). Genre and the pragmatic concept of background knowledge. In A. Freedman & P. Medway (eds.) (1994), *Genre and the New Rhetoric*, pp. 155-178. London: Taylor & Francis.

Google. (n.d.). Retrieved on January, 20, 2003 from http://www.google.com/

Gordon, M. & Pathak, P. (1999). Finding information on the world wide web: The retrieval effectiveness of search engines. *Information Processing & Management*, 35(2), 141 – 180.

Greenberg, J. (2004). Metadata extraction and harvesting: a comparison of two automatic metadata generation applications. *Journal of Internet Cataloging*, 6(4): 59-82.

Greisdorf, H. & Spink, A. (2001). Median measure: an approach to IR systems evaluation. *Information Processing & Management*, 37(6), 843-857.

Haas, S. & Grams, E. (2000). Readers, authors, and page structure: A discussion of four questions arising from a content analysis of Web pages. *Journal of the American Society for Information Science*, 51, 2, 181-192.

Haas, S. & Grams, E. (1998a). Page and link classifications: Connecting diverse resources. *Proceedings of Digital Libraries '98 – Third ACM Conference on Digital Libraries.* 99-107.

Haas, S. & Grams, E. (1998b). A link taxonomy for web pages. *Proceedings of the 61st American Society for Information Science Annual Meeting.* 485-495.

Haas, S. & Hert, C. (2000) Terminology development and organization in multi-community environments: The case of statistical information. *Proceedings of the 11th American Society for Information Science & Technology SIG/CR Classification Workshop,* 51-72.

Harre, R. (1981). Philosophical aspects of the macro-micro problem. In K. Knorr-Cetina and A.V. Cicourel (eds.) *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies,* Boston, MA: Routledge and Kegan Paul, pp. 1-47.

Herring, S., Scheidt, L., Bonus, S. & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the 38nd Annual Hawaii International Conference on Systems Sciences, Digital Media Track,* IEEE Computer Society Press.

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science, 46,* 400–425.

Hjørland, B. (2002). Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology*, 53(4), 257–270.

Janes, J. (1991a). Relevance judgments and the incremental presentation of document representations. *Information Processing & Management,* 27(6), 629–646.

Janes, J. (1991b). The binary nature of continuous relevance judgments: a study of users' perceptions. *Journal of the American Society for Information Science and Technology*, 42(10), 754–756.

Janes, J. (1993). On the distribution of relevance judgments. *Proceedings of the American Society for Information Science*, 104-114.

Jonassen, D.H., Beissner, K. & Yacci, M. (1993) *Structural Knowledge - Techniques for Representing, Conveying, and Aquiring Structural Knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 45-52.

Karlgren, J., Bretan, I., Dewe, J., Hallberg, A., & Wolkert, N. (1998). Iterative information retrieval using fast clustering and usage-specific genres. *Eighth DELOS Workshop – User Interface in Digital Libraries,* 85-92.

Kekäläinen, J. & Järvelin, K. (2002a). Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. In: Bruce, H., Fidel, R., Ingwersen, P. & Vakkari, P., (eds.) *Emerging Frameworks and Methods, Seattle, 2002,* 253-270.

Kekäläinen, J. & Järvelin, K. (2002b). Using graded relevance assessments in IR evaluation. *Journal of the American Society for Information Science and Technology,* 53(13), 1120-1129.

Kelly, D. (2004). *Understanding implicit feedback and document preference: A naturalistic user study*. Ph.D. Dissertation, Rutgers University.

Kemp, D. (1974). Relevance, pertinence and information system. *Information Storage and Retrieval, 10(2),* 37–47.

Kennedy, G.A. (1991). *Aristotle on Rhetoric: A theory of civic discourse,* New York: Oxford University Press

Kennedy, A. & Shepherd, M. (2005). Automatic identifcation of home pages on the web. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences,* IEEE Computer Society Press. Retrieved on January 27, 2005 from http://www.cs.dal.ca/~shepherd/pubs/home_pages.pdf

Kent A., Belzer, J., Kurfeest, M., Dym, E., Shirey, D. & Bose, A. (1967). Relevance predictability in information retrieval systems. *Methods of Information in Medicine,* 6(2), 45–51.

Kraft, R. & Stata., R, (2003). Finding buying guides with a web carnivore. *Proceedings of the First Latin American World Wide Web Conference (LA-WEB 2003),* 84-92.

Krishnamurthy, S. (2002). The multidimensionality of blog conversations: the virtual enactment of September 11. In Maastricht, The Netherlands: *Internet Research 3.0* conference.

Kwasnik, B. (1999). The role of classification in knowledge representation and discovery. *Library Trends,* 48 (1), 22-47.

Kwasnik, B., Crowston, K., Nilan, M. & Roussinov, D. (2000). Identifying document genre to improve Web search effectiveness. *Bulletin of the American Society for Information Science*, 27(2), 23-26.

Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind.* Chicago: University of Chicago Press.

Lan, W. (2002). *From document clues to descriptive metadata: document characteristics used by graduate students in judging the usefulness of web documents.* Unpublished doctoral dissertation. The University of North Carolina, Chapel Hill, NC.

Lancaster, J. W. (1986). *Vocabulary Control for Information Retrieval, 2nd ed.,* Arlington, Va.: Information Resources Press.

Lancaster, F.W. & Warner, A.J. (1993). *Information Retrieval Today*. Arlington: Information Resources Press.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, (33), 159-174.

Lee, D. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through The BNC jungle. *Language, Learning & Technology,* 5 (3), 37-72.

Lee, Y, & Myaeng, S. (2002). Text genre classification with genre-revealing and subject-revealing features. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 145-150.

Lesk, M., & Salton, G. (1968). Relevance assessments and retrieval system evaluation. *Information Storage and Retrieval, 4(3),* 343–359.

Macskassy S., Banerjee A., Davison B., & Hirsh, H. (1998) Human performance on clustering web pages: a preliminary study. *Proceedins of The Fourth International Conference on Knowledge Discovery and Data Mining*, 264-268.

Maglaughlin, K., & Sonnenwald, D. (2002). User perspectives on relevance criteria: a comparison among relevant, partially relevant, and not-relevant judgments. *Journal Of The American Society For Information Science And Technology,* 53(5), 327–342.

Mani, I., House, D., Klein, G., Hirschman, L., Obrst, L., Firmin, T., Chrzanowski, M., & Sundheim, B. (1999). *The TIPSTER SUMMAC Text Summarization Evaluation*. The MITRE Corporation Technical Report MTR 98W0000138, McLean, Virginia.

Marcus, R., Kugel, P., & Benenfeld, A. (1978). Catalog information and text as indicators of relevance. *Journal of the American Society for Information Science,* 29(1), 15–30.

McLellan, P., Tombros, A., Jose, J., Ounis, I., & Whitehead, M. (2001). Evaluating summarisation technologies: a task-oriented approach. *Proceedings of the 1st International Workshop on New Developments in Digital Libraries (NDDL-2001), International Conference on Enterprise Information Systems (ICEIS 2001)*, 99-112.

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, vol. 70, pp. 151-167. Reprinted in A. Freedman & P. Medway (eds.) (1994), *Genre and the New Rhetoric*, pp. 23-42. London: Taylor & Francis.

Miller, C. (1994). The cultural basis of genre. In A. Freedman & P. Medway (eds.) (1994), *Genre and the New Rhetoric*, pp. 67-78. London: Taylor & Francis.

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9), 810-832.

Murphy, G. & Lassaline, M. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories*, pp. 93-131. Cambridge, MA: MIT Press.

Nilan, M., Pomerantz, J. & Paling, S. (2001). Genres from the bottom up: What has the Web brought us? *Proceedings of the American Society for Information Science and Technology Annual Meeting*, 330-339.

Orlikowski, W. & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly*, (39) 4, 541-574.

Paltridge, B. (1995). Working with genre: A pragmatic perspective. *Journal of Pragmatics, 23*, 393-406.

Park, T. (1992). The nature of relevance in information retrieval: An empirical study. Unpublished doctoral dissertation, School of Library and Information Science, Indiana University, Bloomington, IN.

Petersen, T. (1994). Art and Architecture Thesaurus. New York: Oxford.

Prime-Claverie, C., Biegbeder, M., & Lafouge, T. (2004). Transposition of the cocitation method with a view to classifying web pages. *Journal of the American Society for Information Science and Technology*, 55 (14), 1282-1289.

Rehm, G. (2002). Towards automatic Web genre identification. *Proceedings of the 35nd Annual Hawaii International Conference on Systems Sciences, Digital Documents Track*, IEEE Computer Society Press. Retrieved on September 17, 2003 from http://www.computer.org/proceedings/hicss/1435/volume4/1435toc.htm

Rees, A. & Saracevic, T. (1963). Conceptual analysis of questions in information retrieval systems. *Proceedings of the American Documentation Institute*, Part 2, 175-177.

Rosch, E. (1978). Principles of categorisation. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorisation.* pp. 27-48. Hillside, NJ: Lawrence Erlbaum.

Rose, D. & Levinson, D. (2004). Understanding user goals in web search. *Proceeding of the WWW 2004 Conference*, 13-19.

Ross, S. (2001). Cyberculture, cultural asset management, and ethnohistory – preserving the process and understanding the past**.** Keynote Address at: *Preserving the present for the future - Strategies for the Internet*, Blixensalen, The Royal Library, Copenhagen June 18-19 2001. Retrieved on January 5, 2005 from: http://www.deflink.dk/upload/doc_filer/doc_alle/740_sross_cyberculture_rev2.pdf

Rosso, M. (2003). Web information retrieval by genre literature review. Unpublished manuscript.

Rosso, M. (2002). User recognition of web genres. Unpublished manuscript.

Roussinov, D., Crowston, K., Nilan, M., Kwasnik, B., Cai, J. & Liu, X. (2001). Genre based navigation on the web. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Digital Documents Track*, IEEE Computer Society Press. Retrieved on September 17, 2003 from http://www.computer.org/proceedings/hicss/0981/volume%204/0981toc.htm

Rugg, G. & McGeorge, P. (2003). Sorting and categorization. *The Encyclopedia of Library and Information Science*, volume 69, pp. 2699-2707 Marcel Dekker, Inc, New York.

Rugg, G. & McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2), 80-93.

Ryan, T., Field, R., & Olfman, L. (2002). Homepage genre dimensionality. *Proceedings of the Eighth Americas Conference on Information System*, pp. 1116-1128.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24 (5), 513-523.

Salton, G. & Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41 (4), 288-297.

Santini, M. (2005). Genres in formation? An exploratory study of web pages using cluster analysis. *Proceedings of the 8th Annual Colloquium for the UK Special Interest Group forComputational Linguistics (CLUK 05)*, University of Manchester (UK).

Saracevic, T. (1969). Comparative effects of titles, abstracts and full texts on relevance judgments. *Proceedings of the American Society for Information Science*, 293–299.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, 321–343.

Saracevic, T. (1991). Individual differences in organizing, searching, and retrieving information. *Proceedings of the 54th American Society for Information Science Annual Meeting*, 82-86.

Schamber, L. & Bateman, J. (1996). User criteria in relevance evaluation: toward development of a measurement scale. *Proceedings of the 59th American Society for Information Science Annual Meeting*, 218-225.

Schamber, L. (1994). Relevance and information behavior. In: M. E. Williams (Ed.), *Annual Review of Information Science and Technology*, 29 (pp. 3-48). Medford NJ: Learned Information, Inc.

Schamber, L. (1991). Users' criteria for evaluation in a multimedia environment. *Proceedings of the 54th American Society for Information Science Annual Meeting,* 126–133.

Schamber, L., Eisenberg, M., & Nilan, M. (1990). A re-examination of relevance: toward a dynamic, situational definition Source. *Information Processing & Management*, 26(6), 755-776.

Schryer, C. (2002). Genre and power: a chronotopic analysis. In R. Coe, L. Lingard, and T. Teslenko (Eds.), *The Rhetoric and Ideology of Genre*, (pp. 73-102). London: Taylor & Francis.

Shepherd, M. & Watters, C. (2004). Identifying web genre: hitting a moving target. *World Wide Web 2004 Conference Workshop*. Retrieved on November 29, 2004 from http://www.cs.dal.ca/~watters/www2004WorkShop/pdfs/4.pdf

Shepherd, M. & Watters, C. (1999). The functionality attribute of cybergenres. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, Digital Documents Track*, IEEE Computer Society Press. Retrieved on September 17, 2003 from http://www.computer.org/proceedings/hicss/0001/00012/0001toc.htm

Shepherd, M. & Watters, C. (1998). The evolution of cybergenres. *Proceedings of the 31st Annual Hawaii International Conference on Systems Sciences, Digital Documents Track*, IEEE Computer Society Press. 97-109.

Shepherd, M., Watters, C., Duffy, J., & Kaushik, R. (2001a). Browsing and Keyword-Based Profiles: A Cautionary Tale. *Proceedings of the 34th Annual Hawaii International Conference on Systems Sciences, Digital Documents Track*, IEEE Computer Society Press. Retrieved on September 17, 2003 from http://www.computer.org/proceedings/hicss/0981/volume%204/0981toc.htm

Shepherd, M., Watters, C. & Kaushik, R. (2001b). Lessons from reading e-news for browsing the Web: the roles of genre and task. *Proceedings of the 64th American Society for Information Science and Technology Annual Meeting.* 256-267.

Shou, X. Sanderson, M. & Tuffs, N. (2003). The relationship of word error rate to document ranking. *Proceedings of the AAAI Spring Symposium Intelligent Multimedia Knowledge Management Workshop*, Technical Report SS-03-08, 28-33.

Sidler, M. (2002). Web research and genres in online databases: when the glossy page disappears. *Computer and Composition*, (19), 57-70.

Smith, E. & Medin, D. (1981). *Categories and Concepts*. Cambridge, Mass.: Harvard University Press.

Solomon, P. (1992). User-based methods for classification development. *Advances in Classification Research*, 2, 163-170.

Sormunen, E. (2002). Liberal relevance criteria of TREC - counting on negligible documents? In: Beaulieu, M. et al. (Eds): *Proceedings of the Twenty-Fifth Annual ACM SIGIR Conference on Research and Development in Information Retrieval.*

Spink, A. (2002). A user-centered approach to evaluating human interaction with web search engines - an exploratory study. *Information Processing & Management*, 38(3), 401–426.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: Examining different regions of relevance. *Information Processing & Management*, 34(5), 599–621.

Spinuzzi, C. (2004). Four ways to investigate assemblages of texts: genre sets, systems, repertoires, and ecologies. *Proceedings of the 22th Annual International Conference on Design of Communication*, 110-116.

Spinuzzi, C. (2002) Modeling genre ecologies. *Proceedings of the 20th Annual International Conference on Computer Documentation*, 200-207.

Spinuzzi, C. & Zachry, M. (2000). Genre ecologies: an open-system approach to understanding and constructing documentation. *Journal of Computer Documentation*, 24(3), 169-181.

Spinuzzi, C. (1999a) Designing for lifeworlds: genre and activity in information system design and evaluation. Unpublished dissertation. Iowa State University.

Spinuzzi, C. (1999b). Grappling with distributed usability: a cultural-historical examination of documentation genres over four decades. *Proceedings of the 17th annual international conference on Computer documentation*, 16-21.

Steen, G. (1999). Genres of discourse and the definition of literature. *Discourse Processes, 28*, 109-120.

Stein, B. & Meyer zu Eissen, S. (2004). Genre classification of web pages. *Proceedings of the 27th German Conference on Artificial Intelligence (KI-2004),*

Su, L. (2003). A comprehensive and systematic model of user evaluation of web search engines: I. Theory and background. *Journal Of The American Society For Information Science And Technology*, 54(13), 1175–1192.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.

Tang, R., & Solomon, P. (1998). Toward an understanding of the dynamics of relevance judgment: An analysis of one person's search behavior. *Information Processing & Management*, 34(2/3), 237–256.

Taylor, R. (1968). Question-negotiation and information seeking. *College and Research Libraries*, 29(3), 178-194.

Tombros, A. (1997). Reflecting User Information Needs Through Query Biased Summaries. *MSc Thesis, Technical Report (TR-1997-35) of the Department of Computing Science at the University of Glasgow*.

Toms, E., Campbell, D., & Blades, R. (1999). Does genre define the shape of information: the role of form and function in user interaction with digital documents. *Proceedings of the 62th American Society for Information Science Annual Meeting*, 693-704.

Tuffs, N. (2002*). User's perceptions of relevance using query biased-summaries to judge websites and transcribed spoken documents retrieved by two Internet search engines: Google and SpeechBot*. Unpublished masters thesis. University of Sheffield, UK. Retrieved on December 30, 2004 from: http://dagda.shef.ac.uk/dissertations/2001-02/External/Tuffs_Nick_MScIS.pdf

Upchurch, L., Rugg, G. & Kitchenham, B. (2001). Using card sorts to elicit web page quality attributes. *IEEE Software*, 18(4), 84-89.

van Dijk, T. (1985). Introduction: the common roots of the studies of literature and discourse. In T. A. van Dijk (ed.), *Discourse and Literature*, pp. 1-9. Amsterdam: John Benjamins.

Vaughn, M., & Dillon, A. (1998). The role of genre in shaping our understanding of digital documents. *Proceedings of the 61st American Society for Information Science Annual Meeting*, 559-566.

Wang, P., & White, M. (1995). Document use during a research project: a longitudinal study. *Proceedings of the 58th American Society for Information Science Annual Meeting*, 181–188.

White, R., Jose, J., & Ruthven, I. (2003a). A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Information Processing & Management*, 39(5), 707–733.

White, R., Jose, J., & Ruthven, I. (2003b). A granular approach to web search result presentation. *Proceedings of the 9th IFIP TC13 International Conference on Human-Computer Interaction.* Zurich.

Williams, A. (2004). The documentation of quality engineering: applying use cases to drive change in software engineering models. *Proceedings of the 22th annual international conference on Design of communication*, 4-13.

Williams, A. (2003a). Examining the use case as genre in software development and documentation. *Proceedings of the 21st annual international conference on Documentation,* 12-19.

Williams, A. (2003b). Assessing genre as rhetorical performance in software design. *Proceedings of the Professional Communication Conference,* 279-284.

Wittgenstein, L. (1953). *Philosophical Investigations.* New York: Macmillan.

Yates, J., & Orlikowski, W. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review,* vol. 17, no. 2, pp. 299-326.