A Framework for Effective Annotation of Information from Closed Captions Using Ontologies[1]

Latifur Khan
Department of Computer Science
University of Texas at Dallas
Richardson, TX 75083-0688
lkhan@utdallas.edu

Dennis McLeod
Department of Computer Science
University of Southern California
Los Angeles, CA 90088
mcleod@usc.edu

Eduard Hovy
Information Sciences Institute
University of Southern California
Marina del Rey, CA 90292
hovy@isi.edu

ABSTRACT

To improve the accuracy in terms of precision and recall of an audio information retrieval system we have created a domain-specific ontology (a collection of key concepts and their interrelationships), as well as a novel, pruning algorithm. Given the shortcomings of keyword-based techniques, we have opted to employ a concept-based technique utilizing this ontology. Achieving high precision and high recall is the key problem in the retrieval of audio information. In traditional approaches, high recall is typically achieved at the expense of low precision, and vice versa. Through the use of a domain-specific ontology appropriate concepts can be identified during metadata generation (description of audio) or query generation, thus improving precision.

When irrelevant concepts are associated with queries or documents there is a loss of precision. On the other side of the coin, if relevant concepts are discarded, a loss of recall will ensue. In conjunction with the use of a domain specific ontology we have thus proposed a novel, automatic pruning algorithm which prunes as many irrelevant concepts as possible during any case of description and identification of documents, and query generation. Through the association of concepts in the ontology, through techniques of correlation, this pruning algorithm presents a method for the selection of concepts in the query generation. To improve recall, A controlled and correct query expansion mechanism is proposed for the improvement of recall, thus guaranteeing that precision will not be lost.

We have constructed a demonstration prototype with a focus on audio data. Experimentally and analytically we have shown that our model, compared to keyword search, achieves a significantly higher degree of precision and recall. Furthermore, the techniques employed can be applied to the problem of information selection in all media types.

Keywords: Metadata, Ontology, Audio, and SQL.

[1] [1] This research has been funded [or funded in part] by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152.

1 Introduction

New search and retrieval methods are more and more in demand as the amount of useful multimedia information being created (e.g., on the web) continues to swell. Most challenging is the large quantity of non-textual information, such as audio, video, and images, as well as more familiar textual information must be accommodated. The fact that users can be easily overwhelmed by the amount of information available via electronic means is a key problem. Network bandwidth is often wasted through the transfer of irrelevant information in the form of documents (e.g. text, audio, video) retrieved by an information retrieval system and which are of no use to the user. This condition, which can be a source of frustration to the user, is a result of inaccuracies in the representation of the documents in the database, as well as confusion and imprecision in user queries, conditions that result from the fact that users are frequently unable to express their needs efficiently and accurately. These factors contribute to the loss of information and to the provision of irrelevant information. This means that the key problem to be addressed in information selection is the development of a search mechanism to guarantee the delivery of a minimum of irrelevant information (high precision), as well as insuring that relevant information is not overlooked (high recall).

Keyboard based techniques comprise the traditional solution to the problem of recall and precision in information retrieval. Documents are retrieved if they contain a keyword or some combination of keywords specified by the user. But it will be the case that many documents contain the desired semantic information, even though they do not contain the user specified keywords. However, this limitation can be addressed through the use of a query expansion mechanism. Through such a mechanism, additional search terms are added to the original query based on the statistical co-occurrence of terms [16]. Under these circumstances recall will be increased, but generally at the expense of deteriorating precision [14, 20]. We have designed and implemented a concept-based model

using ontologies in order to overcome the shortcomings of keyword-based technique in responding to information selection requests [10]. This model, which employs a domain dependent ontology, is presented in this paper. An ontology is a collection of concepts and their interrelationships which can collectively provide an abstract view of an application domain [4, 7].

The use of an ontology-based model presents two key problems: one is the extraction of the semantic concepts from the keywords and the other is document indexing. With regard to the extraction of semantic concepts, the first problem, the key issue is to identify appropriate concepts which describe and identify documents on the one hand, and on the other, the language employed in user requests. Here, it is important to make sure that irrelevant concepts will not be associated and matched, and that relevant concepts will not be discarded. High precision and high recall, in other words, will be preserved during concept selection for documents or user requests. To address the first problem, in this paper we propose an automatic mechanism for the selection of these concepts from the description of documents/user requests. Through this mechanism irrelevant concepts will be pruned while allowing relevant concepts to become associated with documents/user requests. In addition, a novel, scalable disambiguation algorithm for concept selection from documents using domain specific ontology is presented.

Document indexing is the second problem. To address this, one can use a vector space model of concepts or a richer and more precise model that will employ ontology. We adopt the latter approach. The vector space model does not work well for short queries, hence our choice of a model which uses an ontology. Furthermore, a recent survey on web search engines suggests that average length of user request is 2.2 keywords [3]. To deal with this characteristic, we have developed a concept-based model, which uses domain dependent ontologies for responding to information selection requests. We also propose, in order to improve precision, an automatic query expansion mechanism which deals with user requests expressed in natural language. This automatic expansion mechanism generates database

queries by allowing only appropriate and relevant expansion (please see [11, 23, 24] for more details). Intuition suggests that in order to improve recall during the phase of query expansion, only controlled and correct expansion should be employed, guaranteeing that precision will not be degraded as a result of this process. Further, for the disambiguation of concepts, only the most appropriate concepts must be selected with reference to documents or to user requests. This can be achieved by taking into account the encoded knowledge in the ontology.

Through the exploration and provision of a specific solution to the problem of retrieving audio information, the effectiveness of our disambiguation model can be shown. Tasks entailed by our approach to solving the problem of effective selection/retrieval of audio information include metadata generation (description of audio), and the consequent selection of audio information in response to a query. We have shown that ontologies can be fruitfully employed in connection with facilitating metadata generation. This will require carrying out content extraction by relying on speech recognition technology that converts speech to text. To facilitate information selection requests, we can, after generating transcripts, deploy our ontology-based model. At present, an experimental prototype of the model has been developed and implemented. Our working ontology, at present, has around 7,000 concepts for the sports news domain, with 2,481 audio clips/objects in the database. We use CNN broadcast sports and Fox Sports audio for sample content, along with closed captions. Through the use of our disambiguation algorithm, these associated closed captions are connected with the ontology. Through the study of what percentage of the audio objects selected are associated with relevant concepts we can assess the performance of our disambiguation algorithm. We have observed that through the use of this algorithm 90.5% of the objects extrapolated from closed captions successfully associate with concepts of ontologies, while only 9.5% of the objects fail to associate. Up to 76.9% of the objects

among the 90.5% are associated with relevant concepts (pure); in other cases, objects are associated with relevant and irrelevant concepts (mixed).

By taking the most widely used vector space model as representative of keyword search we can illustrate the power of ontology-based over keyword-based search techniques. For comparison metrics we have used measures of precision and recall, and an F score that is the harmonic mean of precision and recall. Nine sample queries were run based on the categories of broader query (generic), narrow query (specific), and context query formulation. On average our ontology outperforms keyword-based technique. For broader and context queries, the result is more pronounced than in cases of narrow queries.

The structure of the paper will be as follows. In Section 2, we will review related work. In Section 3, we will introduce the research context in terms of the information media used (i.e., audio) and some related issues that arise in this context. In Section 4, we introduce our domain dependent ontology. In Section 5, we present our heuristic-based concept selection mechanism, including the disambiguation algorithm that allows us to choose appropriate concepts for the audio information unit. We also discuss metadata management issues. In Section 6 we give a detailed description of the prototype of our system, and provide data showing how our ontology-based model compares with traditional keyword-based search technique. Finally, in Section 7 we present our conclusions and plans for future work.

It is important to note that part of this research has appeared in the VLDB journal [29]. Reflecting a more complete understanding of this research, some part of this paper will overlap with the previously published work (in particular, Section 2, part of Section 3, part of Section 4, part of Section 6, and part of Section 7).

2 Related Work

Historically ontologies have been employed to achieve better precision and recall in text retrieval systems [9]. Here, attempts have taken two directions, query expansion through the use of semantically related terms, and the use of conceptual distance measures, as in our model. Among attempts using semantically related terms, query expansion with a generic ontology, WordNet [13], has been shown to be potentially relevant to enhanced recall, as it permits matching a query to relevant documents that do not contain any of the original query terms. Voorhees [18] manually expands 50 queries over a TREC-1 collection using WordNet, and observes that expansion was useful for short, incomplete queries, but not promising for complete topic statements. Further, for short queries, automatic expansion is not trivial; it may degrade rather than enhance retrieval performance. This is because WordNet is too incomplete to model a domain sufficiently. Furthermore, for short queries less context is available, which makes the query vague. Therefore, it is difficult to choose appropriate concepts automatically.

The notion of conceptual distance between query and document provides an alternative approach to modeling relevance. Smeaton et al. [16] and Gonzalo et al. [6] focus on managing short and long documents, respectively. Note here that in these approaches queries and document terms are manually disambiguated using WordNet. In our case, query expansion and the selection of concepts, along with the use of the pruning algorithm, is fully automatic.

3 Research Context: Content Extraction

To specify the content of media objects two main approaches have been employed: fully automated content extraction [9] and selected content extraction [19]. In fully automated content extraction, speech is converted to equivalent text (e.g., Informedia). Word-spotting techniques can provide selected content extraction in a manner that will make the content extraction process automatic. Word-spotting is a particular application of automatic speech recognition techniques in which the vocabulary of interest is

relatively small. In our case, vocabularies of concepts from the ontology can be used. Furthermore, content description can be provided in plain text, such as closed captions. However, this manual annotation is labor intensive. For content extraction we rely on closed captions that came with audio object itself from Fox sports and CNN web site in our case (see Section 6).

Definition of an Audio Object

An audio object, by definition and in practice, is composed of a sequence of contiguous segments. Thus, in our model the start time of the first segment and the end time of the last segment of these contiguous segments are used respectively to denote start time and end time of the audio object. Further, in our model, pauses between interior segments are kept intact in order to insure that speech will be intelligible. The formal definition of an audio object indicates that an audio object's description is provided by a set of self-explanatory tags or labels using ontologies. An audio-object O_i is defined by five tuple (id_i, S_i, E_i, V_i, A_i) where Id_i is an object identifier which is unique**,** S_i is the start time, E_i is the end time, V_i (description) is a finite set of tags or labels, i.e., $V_i=\{v_{1i}, v_{2i}, \ldots, v_{ji}, \ldots, v_{ni}\}$ for a particular j where v_{ji} is a tag or label name, and A_i is simply audio recording for that time period. For example, an audio object is defined as {10, 1145.59, 1356.00, {Gretzky Wayne}, *}. Of the information in the five tuple, the first four items (identifier, start time, end time, and description) are called *metadata*.

4 Ontologies

An ontology is a specification of an abstract, simplified view of the world that we wish to represent for some purpose [5, 8]. Therefore, an ontology defines a set of representational terms that we call *concepts*. Interrelationships among these concepts describe a target world. An ontology can be constructed in two ways, domain dependent and generic. CYC [12], WordNet [13], or Sensus [17] are examples of generic ontologies. For our purposes, we choose a domain dependent ontology. First, this is because a domain dependent ontology provides concepts in a fine grain, while generic ontologies provide concepts in

coarser grain. Second, a generic ontology provides a large number of concepts that may contribute to a larger speech recognition error.

Figure 1 (A) shows an example ontology for sports news, and Figure 1(B) shows in Ontology Web Language (OWL) [29]. Such an ontology is usually obtained from generic sports terminology and domain experts. Here, we represent our ontology as a directed acyclic graph (DAG). Each node in the DAG represents a concept. In general, each concept in the ontology contains a label name and

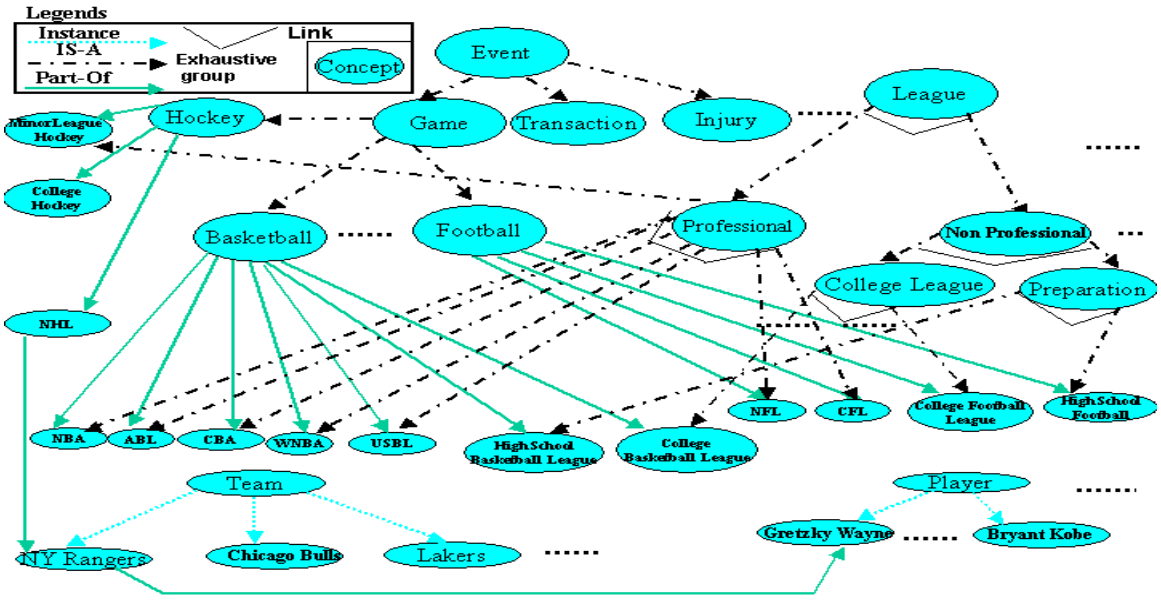

Figure 1 (A). A Small Portion of an Ontology for a Sports Domain

a synonyms list. Note also that this label name is unique in the ontology. Further, this label name is used to serve in associations of concepts with audio objects. The synonyms list of a concept contains vocabulary (a set of keywords) through which the concept can be matched with user requests. Formally, each concept has a synonyms list $(l_1, l_2, l_3, ..., l_i, ..., l_n)$ where user requests are matched with this l_i - an *element* of the list. Note that a keyword may be shared by multiple concepts' synonyms lists. For example, player "Bryant Kobe," "Bryant Mark," "Reeves Bryant" shares common word "Bryant," which may of course create an ambiguity problem.

1.1 Interrelationships

Concepts are interconnected by means of interrelationships. If there is an interrelationship R between concepts C_i and C_j, then there is also a interrelationship R′(R inverse) between concepts C_j and C_i. In Figure 1(A), interrelationships are represented by labeled arcs/links. Three kinds of interrelationships are used to create our ontology: specialization (Is-a), instantiation (Instance-of), and component membership (Part-of). These correspond to key abstraction primitives in typical object-based and semantic data models [2].

IS-A: This interrelationship is used to represent specialization (concept inclusion). A concept represented by C_j is said to be a specialization of the concept represented by C_i if C_j is kind of C_i . For example, “NFL” is a kind of “Professional” league. In other words, “Professional” league is the generalization of “NFL.” In Figure 1 (A), the IS-A interrelationship between C_i and C_j goes from generic concept C_i to specific concept, C_j represented by a broken line. IS-A interrelationship can be further categorized into two types: *exhaustive* and *non-exhaustive*. An exhaustive group consists of a number of IS-A interrelationships between a generalized concept and a set of specialized concepts, and places the generalized concept into a categorical relation with a set of specialized concepts in such a way so that the union of these specialized concepts is equal to the generalized concept. For example, “Professional” relates to a set of concepts, “NBA”, “ABL”, “CBA”, ..., by exhaustive group (denoted by caps in Figure 1(A)). Further, when a generalized concept is associated with a set of specific concepts by only IS-A interrelationships that fall into the exhaustive group, then this generalized concept will not participate in the metadata generation and SQL query generation explicitly. This is because this generalized concept is entirely partitioned into its specialized concepts through an exhaustive group. We call this generalized concept a *non participant concept (NPC)*. For example, in Figure 1(A) “Professional” concept is NPC. On the other hand, a non-exhaustive group consisting of a set of IS-A

does not exhaustively categorize a generalized concept into a set of specialized concepts. In other words, the union of specialized concepts is not equal to the generalized concept.

Instance-Of: The Instance-Of relationship denotes concept instantiation. If a concept C_j is an example of concept C_i, the interrelationship between them corresponds to an Instance-Of denoted by a dotted line. For example, player ”Wayne Gretzky” is an instance of a concept “Player.” In general, all players and teams are instances of the concepts, “Player” and “Team” respectively.

Part-Of: A concept is represented by C_j is Part-Of a concept represented by C_i if C_i has a C_j (as a part) or C_j is a part of C_i. For example, the concept “NFL” is Part-Of the concept “Football” and player, “Wayne Gretzky” is Part-Of the concept “NY Rangers.”

1.2 Disjunct Concept

When a number of concepts are associated with a parent concept through IS-A interrelationships, it is important to note when these concepts are disjoint, and are referred to as concepts of a disjoint type. When, an object belongs to a generalized concept, and is associated with at most one of the concepts from a set of its specific concepts, these specific concepts will be identified as disjoint concepts. For example, the concepts NBA, CBA, or NFL are associated with the parent concept Professional through association with IS-A, they become disjoint concepts. Hence, any given object's metadata cannot possess more than one such concept of the disjoint type. Hence, any given object’s metadata cannot possess more than one such concept of the disjoint type. For example, when an object’s metadata is the concept “NBA,” it cannot be associated with another disjoint concept, such as “NFL.” It is of note that the property of being disjoint helps to disambiguate concepts for keywords in user request (query) disambiguation. Similarly, concepts “College Football”, and “College Basketball” are disjoint concepts due to their associations with parent concept “College League” through IS-A. Furthermore,

“Professional,” and “Non Professional” are disjoint. Thus, we can say that “NBA,” “CBA,” “ABL,” “College Basketball,” and “College Football,” are disjoint.

When a set of concepts are disjoint and each of these disjoint concepts is associated with other concepts through Instance-of/Part-of interrelationships, each of these disjoint concepts along with its associated concepts through Instance-ofPart-of interrelationships will form a *region*. For example, each of these leagues (“NBA,” “CBA,” “ABL,” “College Basketball,” and “College Football,”) and its team and player form a boundary that form a *region*. During annotation of concepts with an audio object we strive to choose a particular region. This is because an audio object can be associated with only one disjoint-type concept. However, it may be possible that a particular player may for example play in several leagues. In that case, we consider two alternatives. First, we will generate multiple instances of the player in the ontology. In other words, for each league in which the player plays he will be represented by a separate concept. In this manner we are able to preserve the property of disjunction. In this case, each region is simply a sub-tree. Second, we will keep just one node for the player that have two parents (say), two teams. In this case, each region is DAG. With the former approach, maintenance or update will be an issue; inconsistency may arise. With the latter approach, maintenance will be easier; however, precision will be hurt. This is because if the query is requested in terms of a team where this player plays, some of retrieved objects will be related to other team and vice versa. This is because both teams have common child concept and query expansion phase allows to retrieve all associated audio objects to this player regardless of his teams. For example, player "Deion Sanders" plays two teams "Dallas Cowboys" (under NFL region) and Cincinnati Reds (under MLB region). If user request is specified by "Dallas Cowboys" some objects will be retrieved that contain information Cincinnati Reds along with Deion Sanders. For this, we adopt former approach.

Concepts are not disjoint, on the other hand, when they are associated with a parent concept through Instance-Of or Part-Of. In these cases, some of these concepts may serve simultaneously as metadata for an audio object. An example would be the case in which the metadata of an audio object are team “NY Rangers” and player “Gretzky Wayne,” where “Grezky Wayne” is Part-Of “NY Rangers.”

```
<?xml version="1.0"?>

<rdf:RDF
   xmlns:rss="http://purl.org/rss/1.0/"
   xmlns="http://a.com/ontology#"
   xmlns:jms="http://jena.hpl.hp.com/2003/08/jms#"
   xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#"
   xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
 xml:base="http://a.com/ontology">
 <owl:Ontology rdf:about="">
  <owl:imports rdf:resource="http://protege.stanford.edu/plugins/owl/protege"/>
 </owl:Ontology>
 <owl:Class rdf:ID="Professional">
  <owl:disjointWith rdf:resource="#Non-professional"/>
  <rdfs:subClassOf rdf:resource="#League"/>
 </owl:Class>
 <owl:Class rdf:ID="Injury">
  <rdfs:subClassOf>
   <owl:Class rdf:ID="Event"/>
  </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="NHL">
        <PartOf rdf:resource="#Hockey"/>
 </owl:Class>
 <owl:Class rdf:ID="Team"/>
 <owl:Class rdf:ID="Player"/>
 <owl:Class rdf:ID="Basketball">
  <rdfs:subClassOf>
   <owl:Class rdf:about="#Game"/>
  </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="MinorLeagueHockey">
        <PartOf rdf:resource="#Hockey"/>
 </owl:Class>
 <owl:Class rdf:ID="Game">
  <rdfs:subClassOf rdf:resource="#Event"/>
 </owl:Class>
 <owl:ObjectProperty rdf:ID="PartOf">
  <rdfs:range rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
 </owl:ObjectProperty>
```

```
 <owl:Class rdf:ID="CollegeHockey">
        <PartOf rdf:resource="#Hockey"/>
 </owl:Class>
 <owl:Class rdf:ID="Transaction">
  <rdfs:subClassOf rdf:resource="#Event"/>
 </owl:Class>
 <owl:Class rdf:ID="Hockey">
  <rdfs:subClassOf rdf:resource="#Game"/>
 </owl:Class>
 <Player rdf:ID="Wayne_Gretzky">
  <PartOf rdf:resource="#NY_Rangers"/>
 </Player>
 <Player rdf:ID="Cobe_Bryant"/>
 <Team rdf:ID="NY_Rangers">
  <PartOf rdf:resource="#NHL"/>
 </Team>
…….
</rdf:RDF>
```

Figure 1 (B). A Small Portion of an Ontology for a Sports Domain in OWL

In Figure 1(B) we show a part of ontology in ontology representation language, OWL (Ontology Web Language [30]). Concepts like, “Professional, “NBA”, and “Team” are treated as Class in OWL; on the other hand, “NY Rangers”, and “Gretzky Wayne” are treated as instance of some classes. IS-A interrelationship is expressed as subClassOf relationship which is defined in OWL (e.g., <owl:Class rdf:ID="Injury"><rdfs:subClassOf><owl:Class rdf:ID="Event"/></rdfs:subClassOf></owl:Class>). Here Part-Of relationship is defined as an ObjectPropoerty. For example, if concept, “NY Rangers” is a part of concept “NHL” then this Part-Of interrelationship will be stated in the following way: <Team rdf:ID="NY_Rangers"><PartOf rdf:resource="#NHL"/></Team>. In this example, “NY Rangers” is declared as an Instance-Of “Team” (using Team tag/class). Disjoint/Disjunct properties of concepts are expressed as disjointWith property in OWL (e.g., <owl:Class rdf:ID="Professional"><owl:disjointWith rdf:resource="#Non-professional"/><rdfs:subClassOf rdf:resource="#League"/></owl:Class>).

2 Metadata Acquisition and Management of Metadata

Metadata acquisition is the name for the process through which descriptions are provided for audio objects. In this section we first, present the model of the mechanism though which concepts are selected from ontologies to facilitate words to meaning mapping, along with the automatic disambiguation

algorithm. Next, we will present metadata management issues that are raised in association with these operations.

5.1 Concept Selection and Disambiguation Mechanisms

Our model features an automatic disambiguation algorithm [11] for choosing appropriate concepts for a group of keywords, and we propose further refinements along these lines.

For each audio object we need to find the most appropriate concept(s). Recall that using word-spotting or closed-captions we get a set of keywords which appear in a given audio object. Now we need to map these keywords in conceptual space. In other words, we need to extract concepts from keywords. This is because matching between user requests and documents is done in conceptual space rather than through keyword matching. For this, concepts from ontologies will be selected based on matching terms taken from their lists of synonyms with those based on specified keywords. Furthermore, each of these selected concepts will have a score based on a partial or a full match. It is important to note that keywords in the list of synonyms might only be a variant of keywords present in a relevant document. Plural, gerund forms, and past-tense suffixes are examples of syntactic variations which prevent a perfect match between keywords from the list of synonyms and keywords in matching documents. This problem can be partially overcome through replacing these keywords with their respective stems. This is called *stemming* [29]. A stem is the portion of the keyword which is left after the removal of its affixes (i.e., prefixes and suffixes). For example, connect is the stem for the variants: connected, connecting, connection, and connections. For stemming we used the same algorithm employed in WordNet [28].

It is possible that a particular keyword may be associated with more than one concept in the ontology. In other words, association between keyword and concept is one:many, rather than one:one. Therefore, the disambiguation of concepts is required. The basic notion of disambiguation is that a set

of keywords occurring together determine a context for one another, according to which the appropriate senses of the word (its appropriate concept) can be determined. Note, for example, that base, bat, glove may have several interpretations as individual terms, but when taken together, the intent is obviously a reference to baseball. The reference follows from the ability to determine a context for all the terms.

5.2 Disambiguation Methods

Thus, extending and formalizing the idea of context in order to achieve the disambiguation of concepts, we propose an efficient pruning algorithm based on two principles: co-occurrence and semantic closeness. This disambiguation algorithm first strives to disambiguate across several regions using first principle, and then disambiguates within a particular region using the second. The basic procedure is as follows. For automatic disambiguation within an ontology a set of regions representing different concepts can be defined. The concepts, as they appear in a given region, will be mutually disjoint from the concepts of other regions. This becomes the basis for determining a group of appropriate concepts for a given keyword or collection of keywords. In short, after keywords are matched to the concepts of a given ontology, the region within the ontology in which the greatest number of selected concepts occurs is determined. This region, the one containing the largest number of selected concepts, will at the time be used to associate with documents and user requests. The selected concepts of other different regions will be pruned automatically.

A simple example will make this clear. The keyword "Charlotte" for a particular document is associated with two concepts of the ontology "Charlotte Hornets" and "UNC Charlotte." One is in the region encompassing a professional league, the National Basketball Association, (NBA), the other in the region encompassing college basketball. Thus, at various levels of complexity beyond this simplified example, the disambiguation technique used to distinguish between concepts is based on the general idea that any set of keywords occurring together in context will together determine appropriate concepts

for one another, i.e., fall into the same region, in spite of the fact that each individual keyword is multiply ambiguous.

However, since any keyword alone will determine a group of concepts which are both relevant and irrelevant, and which can occur in different regions, we will need to have a way of dealing with the possibility that even within a region selected for annotation a given keyword will match more than one concept. In other words, within a given region multiple ambiguous concepts will have been selected for a particular keyword, necessitating further disambiguation. In order to further prune irrelevant concepts we will need to determine the correlation between concepts selected in a given region. For this, we use the second principle; semantic distance (in the ontology). When concepts are correlated, concepts closely associated will be given greater weight. This association will be based on minimal distance in the ontology and the matching scores of concepts based on the number of keywords they match. Thus, selected concepts which correlate with each other will have a higher score, and a greater probability of being retained than non-correlated concepts. If scores of particular ambiguous concepts fall below a certain *threshold-score*, which will be a minimum score chosen for selected concepts for that particular object, these concepts will be pruned.

For example, the annotated text for a particular audio object might be:

> Lakers keep grooving with 8th straight win. *Kobe Bryant* scores 21 points as the *Lakers* remain perfect on their *eastern* road trip with a 97-89 triumph over the *Nets*. *Bryant* discussed the eight game win streak and his performance in the All Star game.

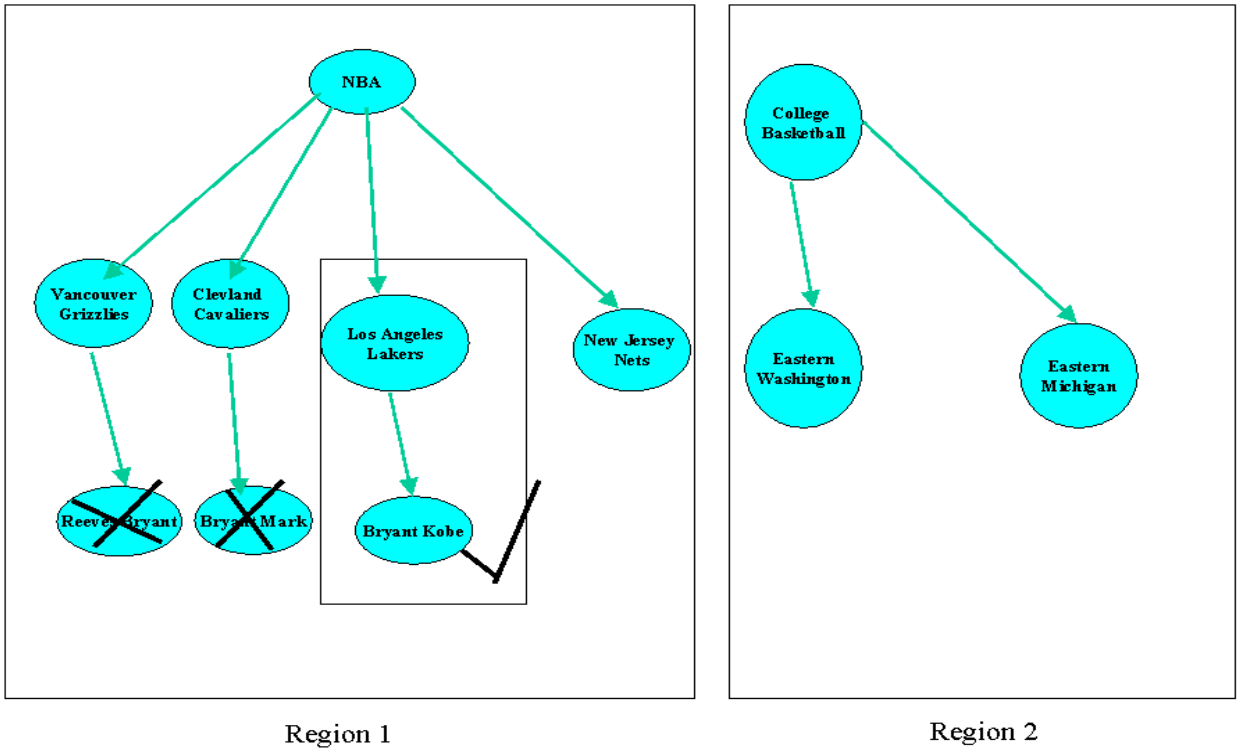

Figure 2. Different Regions of Ontology and Disambiguation of Concepts in Region

The words in italics are the keywords which are associated with the concepts of our ontology. The keywords "Lakers," and "Nets" are associated with the concepts "Los Angeles Lakers" and "New Jersey Nets" respectively. The keyword "Bryant" is associated with the concepts, "Reeves Bryant," "Bryant Mark," and "Bryant Kobe."

It is important to note that all the concepts selected above are found in the region "NBA." However, the keyword "Eastern" is associated with the concepts "Eastern Washington," and "Eastern Michigan" which are not associated with "NBA" but with the region "College Basketball" (see Figure 2). If we now choose only the concepts which appear in the region NBA, which is the region in which the greatest number of concepts occur, the concepts "Eastern Washington" and "Eastern Michigan" will be eliminated, since they are not found in that region. Thus, we keep from among the concepts selected

those which appear in the region NBA in which the greatest number of concepts occur, and prune other selected concepts.

In the selected region, in this case NBA, a keyword such as "Bryant" may be associated with more than one selected concept. This necessitates further disambiguation. We will want to know what other concept qualifies the concepts selected by keyword "Bryant" through correlation. As noted above in the case of the keyword "Bryant" the concepts "Bryant Kobe," "Bryant Mark," and "Reeves Bryant" are all selected. Among these ambiguous concepts, however, only "Bryant Kobe" is correlated with another selected concept, in this case "Los Angeles Lakers." Therefore, "Bryant Kobe" is kept, and the concepts "Bryant Mark," and "Reeves Bryant" are thrown away (see Figure 2—left side).

Thus, we determine the correlation of selected concepts in the region in which the greatest number of keywords have been matched to the audio annotated text, and within that region non-correlated ambiguous concepts are pruned. Finally, the selected concepts for this audio object are "New Jersey Nets," "Los Angeles Lakers," and "Bryant Kobe."

Further, the following example illustrates how the disambiguation algorithm discards irrelevant concepts where a particular audio object semantically carries little information about these concept(s). The annotated text for a particular audio object might be:

> After only two games back, *NBA* bad boy *Dennis Rodman* of the *Dallas Mavericks* has been ejected, fined and suspended. *Rodman* has been suspended without pay for one game and fined $10,000 by the *NBA* for his actions during Tuesday night's home loss to the *Milwaukee Bucks*. *Rodman* expressed his dissatisfaction with the suspension Wednesday by challenging *NBA* commissioner *David Stern* to a boxing match.

The concepts chosen for this audio object by our disambiguation algorithm are player "Dennis Rodman," team "Dallas Mavericks," and team "Milwaukee Bucks," all of which belong to the region

"NBA." It is important to note that our disambiguation algorithm chooses relevant concepts correctly, while irrelevant concepts are automatically pruned (e.g., the concept “Boxing”). If, as in this case, the user request embodies the term boxing, our ontology-based model will not retrieve this object. By contrast, this object will be retrieved when keyword-based technique is employed, with a consequent loss of precision, even though the concept of boxing is not part of what is required to provide conceptual closure in this query.

We have implemented the above idea using score-based techniques. To illustrate this technique we first define some terms, and then present our score-based algorithm.

5.3 Formal Definitions

Each selected concept contains a score based on the number of keywords from the list of synonyms which have been matched with the annotated audio text. Recall that in an ontology each concept (C_i) has a complementary list of synonyms ($l_1, l_2, l_3, ..., l_i, ..., l_n$). Keywords in the annotated text are sought which match each keyword on the element l_j of a concept. The calculation of the score for l_j, which we designate an *Escore*, is based on the number of matched keywords of l_j. The largest of these scores is chosen as the score for this concept, and is designated *Score*. Furthermore, when two concepts are correlated, their scores, called the *Propagated-score*, are inversely related to their position (distance) in the ontology. Let us formally define each of these scores.

Definition 1: Element-score (Escore): The Element-score of an element l_j for a particular concept C_i is the number of keywords of l_j matched with keywords in the annotated text divided by total number of keywords in l_j.

$$Escore_{ij} \equiv \frac{\#of\ \ keywords\ of\ \ l_j\ \ matched}{\| \#\ of\ \ keywords\ in\ l_j \|} \qquad \textbf{(1)}$$

The denominator is used to nullify the effect of the length of l_j on $Escore_{ij}$ and ensures that the final weight is between 0 and 1.

Definition 2: Concept-score (Score): The Concept-score for a concept, C_i is the largest score of all its element-scores. Thus,

$$Score_i = max\ Escore_{ij} \quad where\ 1 \leq j \leq n \qquad (2)$$

Definition 3: Region-score ($Cscore_R$): The Region-score ($Cscore_R$) for a region R is the summation of Concept-score of selected concepts that are belonged to this region. Note that for ambiguous concepts for a particular keyword, their average concept-score is calculated and added to the sum rather than taken as the mere sum of the individual scores.

Definition 4: Semantic distance (SD (C_i , C_j)): *SD (C_i , C_j)* between concepts C_i and C_j is defined as the shortest path between two concepts, C_i and C_j in the ontology. Note that if concepts are in the same level and no path exists, the semantic distance is infinite. For example, the semantic distance between concepts “NBA” and team “Lakers” is 1 (see Figure 2). This is because the two concepts are directly connected via a Part-Of inter-relationship. Similarly, the semantic distance between “NBA,” and “Bryant Kobe” is 2. The semantic distance between “Los Angeles Lakers,” and “New Jersey Nets” is infinite.

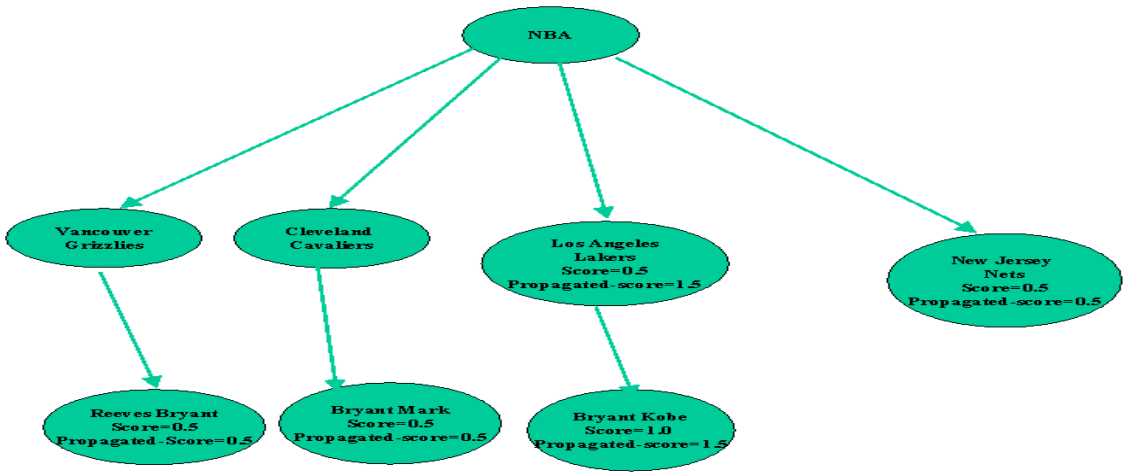

Figure 3. Illustration of Scores and Propagated-scores of Selected Concepts

Definition 5: Propagated-score (S_i): From a set of selected concepts ($(C_j, C_{j+1},...,C_n)$ if a concept, C_i, is correlated with a set of concepts $(C_j, C_{j+1},...,C_n)$, the propagated-score of C_i is its own Score, $Score_i$ plus the scores of each of the rest of selected concepts' $(C_k\ k=j, j+1, ..., n)$ $Score_k$ divided by $SD\ (C_i, C_k)$. Thus,

$$S_i = Score_i + \sum_{k=j}^{k=n} \frac{Score_k}{SD(C_i,C_k)}$$
$$= Score_i + \frac{Score_j}{SD(C_i,C_j)} + \frac{Score_{j+1}}{SD(C_i,C_{j+1})} + ... + \frac{Score_n}{SD(C_i,C_n)} \quad \textbf{(3)}$$

Thus, when two selected concepts are associated/correlated with each other and semantic distance is greater than one, these concepts will have a lower S_i and S_j compared to concepts with the same concept-scores and a semantic distance which is one. This is because for higher semantic distances concepts are correlated in a broader sense. Thus, correlated concepts have a higher S_i than non-correlated concepts. For example, in Figure 3 the values of $Score_i$ for "Los Angeles Lakers" and "Bryant Kobe" are 0.5 and

1.0 respectively. Furthermore, these concepts are correlated with a semantic distance of 1, and their Propagated-scores are 1.5 (0.5 + 1.0/1) and 1.5 (1.0+ 0.5/1) respectively.

Definition 6: S_{max}: For an object, S_{max} is the largest score of all its selected concepts' propagated-score, S_i.

Definition 7: Threshold-score (γ_{Score}): The Threshold-score for an object is a certain fraction of its S_{max}. It is simply determined by the product of S_{max} and a threshold-constant. This threshold-constant can be between 0 and 1.

The pseudo-code for the disambiguation algorithm is as follows:

For each audio object

Find concepts (C_1,C_2,C_3,…,$C_{i,..,}C_m$) that are associated with keywords of annotated text of this audio object

For each region *R*

$Cscore_R = 0$ //initially

//Sum of all selected concepts concept-score for a region, *R*

For each keyword

If a non ambiguous concept C_1 is selected in this region, *R*

//add C_1 score to Region-score

$Cscore_R \equiv Cscore_R + Score_{C1}$

Else If

//ambiguous concepts are selected

Selected ambiguous concepts (C_{k+1},C_{k+2},…,C_{k+r-2},C_{k+r-1},C_{k+r}) are in this region, *R*

//Calculate their average concept-score, $Cscore_{RA}$

$$Cscore_{RA} = \frac{Scorec_{k+1} + Scorec_{k+2} + ... + Scorec_{k+r}}{r}$$

$$Cscore_R = Cscore_R + Cscore_{RA}$$

//End of For Loop for each keyword

//End of For Loop for each region

Choose a region with maximum score, $Cscore_R$

and prune selected concepts in different regions

For this selected region, determine correlation of concepts ($C_i, C_j, C_{j+1}, ..., C_n$) and update their propagated-scores by

$$S_i = Score_i + \sum_{k=j}^{k=n} \frac{Score_k}{SD(C_i, C_k)}$$

$$= \text{Score}_i + \frac{Score_j}{SD(C_i, C_j)} + \frac{Score_{j+1}}{SD(C_i, C_{j+1})} + ... + \frac{Score_n}{SD(C_i, C_n)}$$

//Prune non-correlated ambiguous concepts

Determine maximum score S_{max} among all selected concepts' propagated score S_i for this object

For each ambiguous concept's propagated score S_i

If ($S_i < \gamma_{Score}$ (S_{max} * threshold-constant))

Simply discard this concept which has S_i

Else

Keep this concept

//End of For Loop each ambiguous concept

//End of For Loop for each audio object

Figure 4. Pseudo code for Disambiguation Algorithm

There is a trade-off associated with the selection of value of threshold-constant (γ); γ can be 0, 0.1, 0.2,… For high values of γ, we may lose some relevant concepts and at the same time discard many irrelevant concepts for audio objects. On the other hand, for a lower value of γ, we may keep many irrelevant concepts along with those which are correct. Our goal, for a given audio object, is to keep as many relevant concepts as possible and to throw away the maximum number of irrelevant concepts. By increasing γ, we may discard many ambiguous concepts. In this case, some of those discarded are indeed irrelevant for the object, and by throwing out these concepts better precision can be achieved. This is because in the latter case a given irrelevant object will not be retrieved when the user query is related to one of these discarded concepts.

For the example (given in Figure 2), the concepts "Los Angeles Lakers," "New Jersey Nets," "Bryant Kobe," "Bryant Mark," and "Reeves Bryant," are selected in the selection of region "NBA." S_i, propagated-scores for these concepts are 1.5, 0.5, 1.5, 0.5, 0.5 respectively (see Figure 3). Note that "Los Angeles Lakers," and "Bryant Kobe" are correlated with semantic distance 1 and "Los," and "New" are removed due to the fact that they belong to a stop list of common words. S_{max} is 1.5 here and ambiguous concepts are "Bryant Kobe," "Bryant Mark," and "Reeves Bryant." If we set $\gamma = 0.6$, then the ambiguous concepts "Bryant Mark," and "Reeves Bryant" are discarded since their S_i scores fall below 0.9 ($S_{max} * \gamma = 1.5 * 0.6$). Although S_i for "New Jersey Nets" is 0.5, which falls below the threshold-score, we keep it because it is not an ambiguous concept.

5.4 Characteristics of Disambiguation Algorithm

At this point we present the features of our disambiguation algorithm.

First, through the use of the algorithm it might be possible that a relevant concept may be discarded along with irrelevant ones. This is because a relevant concept may not correlate with other concepts, hence its S_i is low. When relevant concepts are discarded recall will be hurt, because objects with these

concepts will not be retrieved if the user request is framed in terms of these concepts. For example, the annotated text for an audio object is: "*Flyers* fall to *Leafs*. *Eric* scored two goals and the *Leafs* staved off *Flyers'* third-period rally to hang on for a 4-2 victory Wednesday night over the *Philadelphia Flyers*." The concepts "Desjardins Eric," "Lindros Eric", "Philadelphia Flyers", and "Toronto Maple Leafs" are selected. The propagated-scores S_i for these concepts are 1.5, 0.8333, 1.5, and 0.833 respectively. The interrelationships between player "Desjardins Eric," and team "Philadelphia Flyers" and player "Lindros Eric" and team "Toronto Maple Leafs" are Part-Of. If γ=0.6 is chosen as a threshold-constant, among two ambiguous concepts "Lindros Eric" (0.8333 < 1.5*0.6) will be thrown away, and "Desjardins Eric," will be kept. In other words, the relevant concept, "Lindros Eric" will be discarded.

Second, note that if there is no correlation, the algorithm fails to resolve ambiguity. In that case, we keep all the selected concepts. For example, the annotated text for an audio object is: "*Young Tiger* hurlers hoping balance offense." Major league baseball's team "Detroit Tigers" and players "Tiger Dmitri" and "Tiger Eric" are selected. The S_i scores for these concepts are 0.5. Due to a lack of correlations, we cannot throw away irrelevant concepts "Tiger Dmitri" and "Tiger Eric."

Furthermore, due to the incompleteness of the ontology, some irrelevant concepts may be associated with audio objects. For example, the annotated text for an audio object is:

> Team Up exciting part of *NBA* All-Star weekend for commissioner. *NBA* commissioner *David Stern* believes that Team Up, a program that encourages young people to volunteer their time to the community, is the most exciting part of the All-Star weekend. Former players Bob Lanier and *Michael* Cooper agree, and say the program is about making a difference in people's lives.

Among the concepts selected, *NBA* players "Cage Michael," "Curry Michael," "David Kornel," "Dickerson Michael," "Robinson David," "Wingate David," are wrongly selected because our ontology does not contain knowledge about the NBA commissioner.

Third, one important observation is that when a keyword selects one concept we assume that it is unambiguous, although this unambiguous concept may have a low score as a result of not being correlated with other concepts. In the Figure 3, as a case in point, the concept "New Jersey Nets" has S_i=0.5. Further, some of these concepts may not be relevant to audio objects. If the annotated text for an audio object is:

> *Titans* coaches bring game plan to Atlanta. The *Tennessee Titans* fight through the cold of Atlanta and the absence of a bye week to prepare for the SuperBowl against the Rams Sunday. *Titans* quarterback *Steve McNair* believes that the cold *weather* might actually help his turf toe.

Besides, concepts "Mcnair Steve" and "Tennessee Titans," player "Weathers Andre," a concept which is not relevant, is also selected.

Finally, it may be possible that among ambiguous concepts one will simply subsume the other. For example, the annotated text for an audio object is: "*Caps Oates* scores 300th goal; beat Islanders. *Adam Oates* scores his 300th career goal with 5:01 left Monday night, giving the *Washington Capitals* a 3-2 victory over the *New York Islanders*." Concepts "Oates Adam," "Washington Capitals," "York Mike," and "York Rangers" are selected. Note that "York Mike," and "York Rangers" are ambiguous concepts, and the interrelationship between "York Mike" and "York Rangers" is Part-Of. In that case we discard concept, "York Mike." This is because for this audio object, one team "Washington Capitals" has already been selected. Most probably the object conveys information about one team's performance

over the other. It is important to note that if concept "York Mike" is selected with higher S_i, we keep this concept.

Therefore, disambiguation fails to disambiguate concepts when there is little or no context among the concepts selected. This is an extremely rare occurrence. In a case in which it does occur, we keep all selected concepts; where some of them are relevant and some are irrelevant. However, whenever some clue (i.e., additional concepts) is available in almost cases disambiguation discards the irrelevant concepts which have been associated with audio objects. In only a few cases are relevant concepts also discarded.

This way we guarantee that precision will not be hurt in the course of the extraction of concepts from keywords in the phase of document representation. This will contribute a gain over keyword-based search on one side of the coin (i.e., document representation) with the other side of the coin being the querying mechanism (see Chapter 6 of [24]).

3 Experimental Implementation

We have constructed an experimental prototype system which is based upon a client server architecture. The server (a SUN Sparc Ultra 2 model with 188 MBytes of main memory) has an Informix Universal Server (IUS), which is an object relational database system.

For the sample audio content we use CNN broadcast sports audio and Fox Sports. We have written a hunter program in Java that goes to these web sites and downloads all audio and video clips with closed captions. The average size of the closed captions for each clip is 25 words, after removing stop words. These associated closed captions are used to hook with the ontology. As of today, our database has 2,481 audio clips. The usual duration of a clip is not more than 5 minutes in length. Wav and ram are used for media format.

Currently, our working ontology has around 7,000 concepts for the sports domain (see Table 1). For fast retrieval, we load the upper level concepts of the ontology in main memory, while leaf concepts are retrieved on a demand basis. Hashing is also used to increase the speed of retrieval.

Table 1 Parameters Used for Experimental Results

Media support	Wav and Ram
Total # of clips	2,481
Maximum length of clip	5 min
Average size of closed caption for a clip after removing stop words	25 words
Total # of concepts in ontologies	7,000
Average # of concepts associated with an object	4.47

Our prototype system has five components: database, metadata generator, selector, player, and user interface. The selection of concepts and the use of the disambiguation algorithm takes place in the metadata generator module, where the algorithm chooses appropriate concepts for audio clips from their closed captions (shown by 1 in Figure 5; discussed in Section 5). Only the concepts, along with their URLs as metadata, are stored in the database. The database does not contain any audio data. It contains URLs that facilitate downloading audio data on demand from the source in which it is stored. The User Interface handles user requests expressed in the form of natural language, and dispatches these to the selector (shown by 2 in Figure 5). The selector, using the pruning module, chooses relevant concepts and discards those which are irrelevant (discussed in Section 5.2). From the concepts selected, the selector generates database queries in SQL, using an expansion module with possible optimizations, and then submits these to the database (shown by 3 in Figure 5; discussed in Paper [24]). The URLs of relevant clips, with closed captions, are next displayed in the web browser, with the most recent relevant clip shown first. When the user clicks on a closed caption, the browser will invoke the real player/windows media player (shown by 4 in Figure 5). Note that concept selection, the disambiguation modules for metadata generation, and the pruning and query expansion modules with possible

optimizations for selection are all written in Java. Furthermore, the connectivity between the database and these modules has been achieved through JDBC.

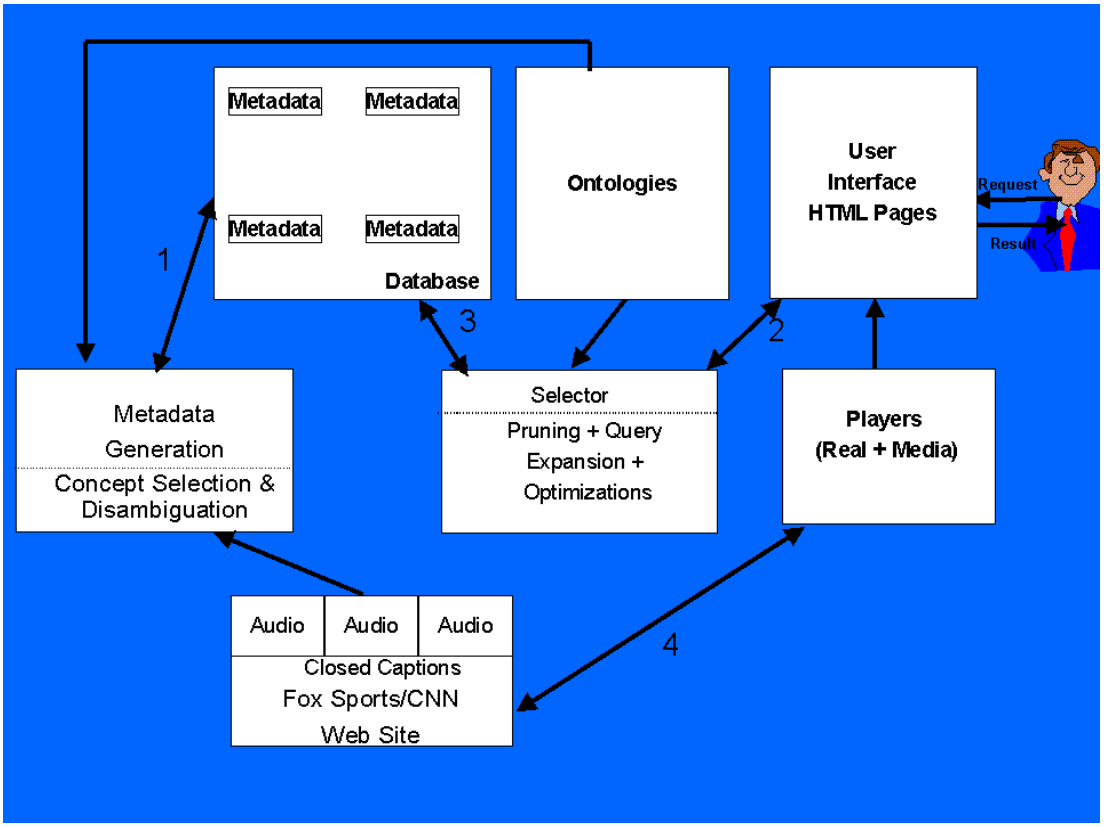

Figure 5. Components of Our Prototype System

6.1 Interface

To enter a query into our system with basic form, users are merely required to type in a few descriptive keywords and to hit "search." Our interface will then provide a set of html pages that contain the hyperlink (absolute URL) of relevant audio objects, along with closed captions (see Figure 6). Each html contains at most 20 audio objects. When a user clicks on a certain closed caption, a real player or windows media player will be invoked, depending on the medium used, and the clip will start to play.

Our interface automatically expands user search using stemming, so when a user types in "boxing" the interface searches for "boxer" as well, and maybe even "box" (discussed in Section 5.1). There is some argument about whether this is good or bad, but it is a necessary reflection of the fact that our matching criterion is from concept to concept rather than keyword to keyword. Furthermore, our disambiguation algorithm always chooses the right concepts from the keywords. Therefore, additional query terms usually do not hurt precision. As discussed in Section 5.1, our Interface removes common words such as "of" and "for" from the query before it starts to search. We ignore these words for two reasons:

- Common words rarely help narrow down a search, and
- Common words slow down searching significantly.

It is important to note that our basic interface will return objects that contain only some of the query terms, not necessarily all of them. For example, a user types "Los Angeles Lakers or Portland Trail Blazers." Our interface will return objects that talk about either "Los Angeles Lakers," or "Portland Trail Blazers," or both. However, if the user types "Los Angeles Lakers and Portland Trail Blazers," the interface will retrieve the previous result set. Thus we do not discriminate between keyword "and" as opposed to keyword "or." Furthermore, when the user types "NBA and NHL," the retrieved object should contain information about either the NBA or the NHL (See Section 6.2 & [24] for more details). This is because these concepts are disjoint. Since our goal is not to achieve a level of understanding comparable to that of natural language, these terms "and," and "or" are added to the list of common words. However, the user can construct conjunctive or difference (not) queries using a few advanced search operators.

6.2 Advanced Search Interface

In conjunctive query the interface returns only objects that include all the query terms. The + operator enforces "and" behavior in our interface (i.e. the user types "Los Angeles Lakers + Portland Trail Blazers"). Sometimes it is helpful to choose words to be excluded from a search. That is, we want all relevant result objects except those conveying certain keywords. We support this "not" functionality with the "-" operator (i.e. the user types "hockey- college hockey." Furthermore, our interface does not discriminate between whether or not the query terms are found in close proximity.

6.3 Results

6.3.1 Effectiveness of Disambiguation Algorithm

We have completed study of the performance of our disambiguation algorithm by considering what percentage of audio objects it can successfully disambiguate. Furthermore, we studied the impact of levels of threshold values on pruning irrelevant concepts associated with audio objects while retaining those which are relevant. For study data, we ran our disambiguation algorithms over the audio clips' closed captions. We then inspected the concepts associated with various audio objects. In Figure 6 the X axis represents the value of threshold, γ, while the Y axis represents the percentage of instances in which objects are annotated with only correct concepts (category I), with wrong concepts (category II), and with no concept at all (category III). In category II, showing wrong concepts, some correct concepts may also be present (mixed).

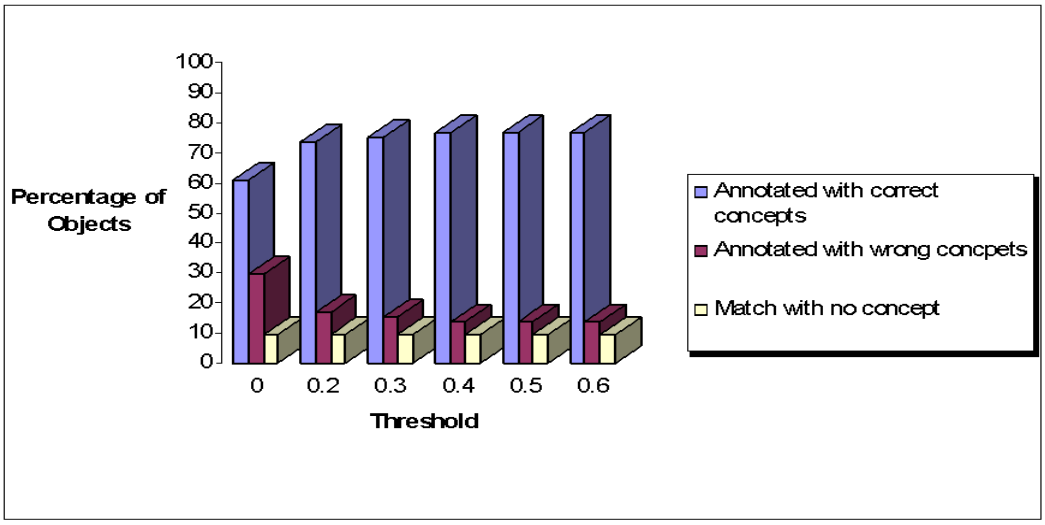

Figure 6. Effect of Threshold on Audio Objects' Associated Concepts

For $\gamma=0$ when the disambiguation algorithm works among different regions, we observed that 9.5% of the objects failed to associate with any concept of the ontology (category III). This is because our ontology is incomplete. For example, an audio object includes reference to a famous hockey player whose career ended ten years ago, and who recently passed away. There is no concept for this player in our ontology, so our algorithm fails to associate a concept with this object. Thus, recall will be hurt. On the other hand, 90.5% of the objects are associated with at least some concepts of the ontology (category I & II). Among these, 60.8% objects are all associated with relevant concepts (category I). In other words, in 60.8% of the cases there is no association with an irrelevant concept. Nor in these cases, have we missed any relevant concept. 29.7% objects are associated with at least one irrelevant concept along with relevant concepts (category II). In this case, precision is hurt due to the annotation of irrelevant concepts. Note that in this case these irrelevant concepts for an audio object are distributed in several regions or a particular region.

With an increasing value of γ, the threshold-constant, ambiguous concepts will be discarded from category II. Furthermore, this threshold-constant strives to resolve ambiguity for an audio object in a particular region, rather than in several regions. Recall that an audio object might be associated with several concepts. From there the S_{max} score is calculated and ambiguous concepts whose propagated-score, S_i, falls below $S_{max} * \gamma$ are simply discarded. Note that S_{max} varies from object to object. Thus, some objects will be rid of irrelevant concepts and will now be associated with correct concepts (category I). However, as emphasized earlier, there is a chance that with the increasing value of γ, for a given audio object, we may lose a relevant concept as we shed those which are irrelevant. Thus, recall will be diminished at the expense of improving precision. For a particular γ in Figure 6, the first, second, and third bars represent category I, II and III respectively. Hence, with γ equal to the values 0.2, 0.3, 0.4, 0.5, and 0.6 respectively, 73.7%, 75%, 76.9%, 76.9%, and 76.9% of the objects are associated with relevant concepts (category I). Further, 16.8%, 15.5%,13.6%, 13.6% and 13.6% of the objects are associated with irrelevant concept(s), along with relevant concept(s), and/or are missing some relevant concepts that are selected a priori (as compared to γ=0). Note also that category III is independent of the increasing value of γ.

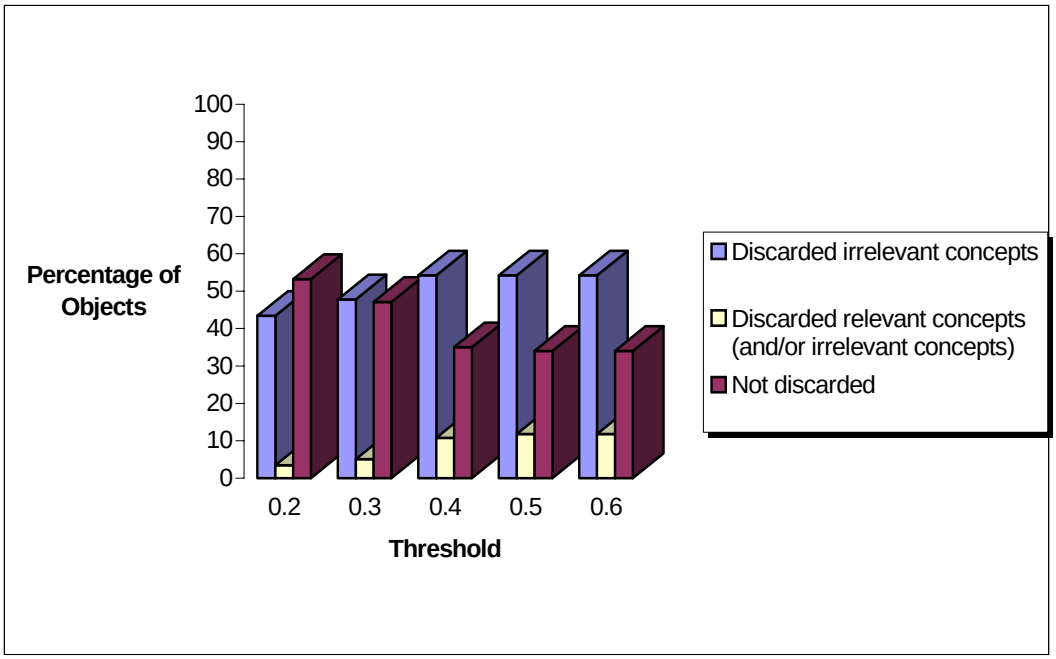

Figure 7. Effect of Threshold on Audio Objects' Irrelevant Concepts (Mixture)

In Figure 7 we show separately the results of our study of the impact of an increasing value of γ for category II. Increasing the value of γ not only leads to the discarding of irrelevant concepts from audio objects but also to the loss of relevant concepts. Here, the X axis represents the threshold value of γ, while the Y axis represents the percentage of objects in which discarded irrelevant concepts and relevant concepts occur for category II. For a particular γ, the first, second, and third bars represent the percentage of objects in which all associated irrelevant concept were discarded, the percentage of objects in which at least one relevant concept was discarded, and the percentage of objects in which no ambiguous concept was discarded out of the 29.7% total objects of category II at γ=0. With γ=0.2, 0.3, 0.4, 0.5, and 0.6, 43.4%, 47.81%, 54.21%, 54.22% and 54.22% of the objects reflect the condition that only irrelevant concepts have been discarded, while only 3.4%, 5.05%, 10.77%, 11.78% and 11.78% of the objects reflect the condition that relevant concepts have been discarded. One important observation

is that with an increasing γ, more objects discarded irrelevant concept(s) as compared to a decreasing number of objects in which correct concepts were missed. For γ=0, 60.8% objects are in category I. With γ=0.2, 0.3, 0.4, 0.5, and 0.6, out of 29.7% objects 12.89% (43.4%* 29.7%), 14.20% (47.81%* 29.7%), 16.10% (54.21%*29.7%), 16.10% (54.22%*29.7%) and 16.10% (54.22%* 29.7%) of the objects are all associated with relevant concepts respectively. These will be added to the 60.8% of the objects associated with relevant concepts at γ=0 and are in category I. Thus, with γ=0.2, 0.3, 0.4, 0.5, and 0.6, 73.69%, 75%, 76.9%, 76.9%, and 76.9% objects are in category I respectively in Figure 6.

Note also, with the increasing threshold-constant, γ curves of categories I, and II (in Figure 6) and all curves in Figure 7 become flat. This is because at γ=0.4 or higher the disambiguation algorithm is unable to throw any new irrelevant/relevant concepts from category II. It is important to note that in our data set the propagated-scores of non-correlated concepts do not fall into this range. Moreover, in our data set, the semantic distance of most of the correlated selected concepts is 1. After the propagation of scores among these concepts their propagated-scores are equal, and they participate in the selection based on the largest scores principle. On the other hand, the propagated-scores of non-correlated concepts are low. When we cannot disambiguate concepts due to unavailability of context (i.e., usually selected concepts' propagated-scores are equal) we simply keep all the concepts, both those which are relevant and those which are not.

Now, the question is what threshold constant should we choose? This will depend on the dataset. However, inability to achieve further progress in distinguishing relevant from irrelevant concepts dictates a limit on increasing the threshold-constant value. This is seen when categories II and III cease changing. With regard to categories II and III the threshold-constant governs the way in which concepts associated with an object are used as metadata to handle user requests. For example, in the above result we increased the threshold-constant 0.1 in each successive iteration. When we observe that categories II

and III have become flat we simply stop and use the value 0.4, which is the maximum value in the set of threshold-constants (γ) from successive iterations. After that the process ceases and no further improvement is possible with the increasing threshold-constants (e.g., 0.5, 0.6, and so forth).

Table 2 Recall/Precision/F score for Two Search Techniques

Types of Queries		Recall		Precision		F score	
		Ontology	Keyword	Ontology	Keyword	Ontology	Keyword
Generic /broader queries	Query 1	90%	11%	98%	95%	94%	20%
	Query 2	95%	15%	89%	90%	92%	26%
	Query 3	87%	30%	100%	82%	93%	44%
Specific/ narrow queries	Query 4	85%	71%	100%	72%	91%	71%
	Query 5	100%	65%	77%	100%	87%	79%
	Query 6	90%	76%	76%	90%	83%	83%
Context queries	Query 7	100%	74%	74%	16%	82%	27%
	Query 8	90%	76%	76%	29%	82%	42%
	Query 9	85%	83%	100%	34%	92%	49%
	Averages	91.3%	55.6%	87.7%	67.5%	88.7%	49%

A set of queries from real users were gathered through this interface they posed. These queries are then grouped into the three categories: broad query, narrow query, and context query. The comparison metrics used for these two search techniques are precision, recall, and F score. For each category, we have sorted queries based on the ascending score of F score metric for keyword-based technique and for each category first three queries results are reported. Hence, worst case result has been reported from our ontology-based model perspective. Thus, results are reported for only nine queries in Table 2. The average *recall* of ontology-based model and keyword technique is 91.3%, and 55.6% respectively. The average *precision* of ontology-based model and keyword technique is 87.7% and 67.5% respectively.

Table 2 also shows the comparative *F score* for these two algorithms. Results show that ontology-based model (average F=88.7%) out performs keyword based technique (average F=49%).

4 Conclusions

We have proposed a potentially powerful and novel approach for the retrieval of audio information. The basis for significant improvement over previous models is the development of an ontology-based model for the generation of metadata for audio, and the selection of audio information in a user customized manner. The ontology we propose can be used to generate information selection requests in database queries, something we have demonstrated in this paper. For a demonstration project we have used a domain of sports news information, but our results can be generalized to fit many additional important content domains including video with closed captions. By providing many different levels of abstraction in a flexible manner with greater accuracy in terms of precision, recall and F score, our ontology-based model has demonstrated its power over keyword based search techniques

We are confident that the fundamental conceptual framework for this project is sound, and its implementation completely feasible from a technical standpoint. The most pressing remaining question relates to the cost of building such domain-specific ontologies and connecting domain data to them automatically. These questions are being addressed by ongoing ontology construction research in the knowledge representation community. Still other tasks remain to be undertaken in future work. These include detailed work on evolving ontologies, extracting highlighted sections of audio, addressing retrieval questions in the video domain, and facilitation of cross-media indexing. It is our goal to build ontology that is easy to update, open and dynamic both algorithmically and structurally for easy construction and modification, and fully capable of adapting to changes and new developments in a domain. Suppose, for example, player "Bryant Kobe" switches from team "Los Angeles Lakers" to team "Portland Trail Blazers." In this case, we need to remove the interrelationship link between

concepts "Bryant Kobe" and "Los Angeles Lakers" and add a new link between the concepts "Bryant Kobe" and "Portland Trail Blazers."

By implication, we would like to address the problem of how to create useful ontology by minimizing the cost of initial creation, while allowing for novel concepts to be added with minimum intervention and delay, and to do so we would like to combine techniques from knowledge representation [25], resource description framework [21, 26], natural language processing, and machine learning.

5 References

[1] B. Arons, "SpeechSkimmer: Interactively Skimming Recorded Speech," in *Proc. of ACM Symposium on User Interface Software and Technology*, pp. 187-196, Nov 1993.

[2] G. Aslan and D. McLeod, "Semantic Heterogeneity Resolution in Federated Database by Metadata Implantation and Stepwise Evolution," *The VLDB Journal, the International Journal on Very Large Databas*es, vol. 18, no. 2, Oct 1999.

[3] R. Baeza and B. Neto, *Modern Information Retrieval*, ACM Press New York, Addison Wesley, 1999.

[4] M. Bunge, *Treatise on basic Philosophy, Ontology I: The Furniture of the World*, vol. 3, Reidel Publishing Co., Boston, 1977.

[5] S. Gibbs, C. Breitender, and D. Tsichritzis, "Data Modeling of Time based Media," in *Proc. of ACM SIGMOD*, pp. 91-102, 1994, Minneapolis, USA.

[6] J. Gonzalo, F. Verdejo, I. Chugur, and J. Cigarran, "Indexing with WordNet Synsets can Improve Text Retrieval," in *Proc. of the Coling-ACL'98 Workshop: Usage of WordNet in Natural Language Processing Systems*, pp. 38-44, August 1998.

[7] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition," *An International Journal of Knowledge Acquisition for Knowledge-based Systems*, vol. 5, no. 2, June 1993.

[8] N. Guarino, C. Masolo, and G. Vetere, "OntoSeek: Content-based Access to the Web," *IEEE Intelligent Systems*, vol. 14, no. 3, pp. 70-80, 1999.

[9] A. G. Hauptmann, "Speech Recognition in the Informedia Digital Video Library: Uses and Limitations," in *Proc. of the Seventh IEEE International Conference on Tools with AI*, Washington, DC, Nov 1995.

[10] L. Khan and D. McLeod, "Audio Structuring and Personalized Retrieval Using Ontologies," in *Proc. of IEEE Advances in Digital Libraries, Library of Congress*, pp. 116-126, Bethesda, MD, May 2000.

[11] L. Khan, and D. McLeod, "Effective Retrieval of Audio Information from Annotated Text Using Ontologies, Proc. of ACM SIGKDD Workshop on Multimedia Data Mining, Boston, MA, pp. 37-45, August 2000.

[12] D. B. Lenat, "Cyc: A Large-scale investment in Knowledge Infrastructure," *Communications of the ACM*, pp. 33-38, vol. 38, no. 11, Nov 1995.

[13] G. Miller, "WordNet: A Lexical Database for English," *Communications of the ACM,*" vol. 38, no. 11, Nov, 1995.

[14] H. J. Peat and P. Willett, "The Limitations of Term Co-occurrence Data for Query Expansion in Document Retrieval Systems," *Journal of ASIS*, vol. 42, no. 5, pp. 378-383, 1991.

[15] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.

[16] A. F. Smeaton and V. Rijsbergen, "The Retrieval Effects of Query Expansion on a Feedback Document Retrieval System," *The Computer Journal*, vol. 26, no. 3 pp. 239-246, 1993.

[17] B. Swartout, R. Patil, K. Knight, and T. Ross, "Toward Distributed Use of Large-Scale Ontologies," in *Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*, Banff, Canada, 1996.

[18] E. Voorhees, "Query Expansion Using Lexical-Semantic Relations," in *Proc. of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 61-69, 1994.

[19] L. D. Wilcox and M. A. Bush, "Training and Search Algorithms for an Interactive Wordspotting System," in *Proc. of ICASSP*, vol. 2, pp. 97-100, San Francisco, CA, 1992.

[20] W. Woods, "Conceptual Indexing: A Better Way to Organize Knowledge," *Technical Report of Sun Microsystems*, 1999.

[21] Using XML: Ontology and Conceptual Knowledge Markup Languages, 1999, http://www.oasis-open.org/cover/xml.html.

[22] J. Hirschberg and B. Grosz, "Intonational Features of Local and Global Discourse," in *Proc. of the Speech and Natural Language Workshop*, pp. 23-26, Harriman, NY, Feb 1991.

[23] L. Khan, D. McLeod and E. Hovy, "Retrieval Effectiveness of Ontology-based Model for Information Selection," the VLDB Journal: The International Journal on Very Large Databases, ACM/Springer-Verlag Publishing, Vol. 13(1): 71-85 (2004).

[24] L. Khan, "Ontology-based Information Selection," Ph.D. Dissertation, Department of Computer Science, University of Southern California, August 2000.

[25] P. Mitra, M. Kersten and G. Wiederhold, "Graph-Oriented Model for Articulation of Ontology Interdependencies" in *Proc. of the 7th International Conference on Extending Database Technology, EDBT 2000.*

[26] Resource Description Framework, http://www.w3.org/RDF/

[27] G. Miller, “Nouns in WordNet: a Lexical Inheritance System," *International Journal of Lexicography*, vol. 3, no. 4, pp. 245-264, 1994.

[28] M. F. Porter, “An Algorithm for Suffix Stripping Program,” Editors J. S. Karen, and P. Willet, *Readings in Information Retrieval*, San Francisco, Morgan Kaufmann, 1997.

[29] OWL Web Ontology Language Overview, http://www.w3.org/TR/2004/REC-owl-features-20040210/