Teaching Machines about Everyday Life

Push Singh, Barbara Barry, and Hugo Liu

20 July 2004

{push, barbara, hugo}@media.mit.edu

Media Lab
Massachusetts Institute of Technology
20 Ames St.
Cambridge, MA 02139
United States

Abstract

In order to build a new breed of software that can deeply understand people and our problems, so that they can help us to solve them, we are developing at the Media Lab a suite of computational tools to give machines the capacity to learn and reason about everyday life—in other words, to give machines 'common sense'. We are building several large-scale commonsense knowledge bases that model broad aspects of the ordinary human world, including descriptions of the kinds of goals people have, the actions we can take and their effects, the kinds of objects that we encounter every day, and so forth, as well as the relationships between such entities. In this article we describe three systems we have built—ConceptNet, LifeNet, and StoryNet—that take unconventional approaches to representing, acquiring, and reasoning with large quantities of commonsense knowledge. Each adopts a different approach: ConceptNet is a large-scale semantic network, LifeNet is a probabilistic graphical model, and StoryNet is a database of story-scripts. We describe the evolution of these three systems, the techniques that underlie their construction and their operation, and conclude with a discussion of how we might combine them into an integrated commonsense reasoning system that uses multiple representations and reasoning methods.

1 Introduction

Can we build a new breed of software with enough 'common sense' to reason in useful ways about ordinary human life? Imagine if your cell phone were smart enough to switch to silent mode when you entered a movie theater but alerted you during the film if a relative were to call from the hospital, but not when your friend called from the pub. Imagine if when you complained "I can't get a good night's sleep", your search engine suggested a mattress sale at a nearby store. Imagine if you entered a child's birthday party into you electronic calendar and it asked, "Do you think a kite would be a good birthday gift?"

Such abilities are beyond today's machines largely because they lack even the most rudimentary understanding of people and the structure of ordinary human life. They don't know anything about, for example:

- the kinds of things we typically do
- the objects we interact with and why
- the consequences of actions in different situations

- the kinds of relationships we have with one another
- the things we like and things we don't like
- the places we find familiar and the things we do there
- the feelings and emotions that motivate us

Instead, our machines today are mindless tools that possess no understanding of why a typical person would need to use them. As a result they are inflexible, unfriendly, and often unnecessarily complicated—they have no ability to adapt to new circumstances, understand the context of their use, or make good guesses at what we wish for them to do, regardless of how obvious it may seem to us.

Can we build machines that can actually understand people, and especially, that can understand our goals and help us achieve them? Our research at the Media Lab is focused on giving machines this kind of understanding. We are interested in the large-scale structure of the human conceptual system, and are working on embodying such structures within our machines, in order to make them more understanding, helpful, and to make our interactions with them more seamless.

We are developing at the Media Lab a suite of computational tools to give machines the capacity to learn and reason about ordinary human life. We are building several large-scale commonsense knowledge bases that model broad aspects of the ordinary human world, including descriptions of the kinds of goals people have, the actions we can take and their effects, the kinds of objects that we encounter every day, and so forth, as well as the relationships between such entities.

In this article we describe three systems we have built—ConceptNet, LifeNet, and StoryNet—that take unconventional approaches to representing, acquiring, and reasoning with large quantities of commonsense knowledge. Each adopts a different approach: ConceptNet is a large-scale semantic network, LifeNet is a probabilistic graphical model, and StoryNet is a database of story-scripts. We describe the evolution of these three systems, the techniques that underlie their construction and their operation, and conclude with a discussion of how we might combine them into an integrated commonsense reasoning system that uses multiple representations and reasoning methods.

1.1 Why is common sense hard?

Compared to other areas of AI, there has been relatively little work on building machines capable of commonsense reasoning about many aspects of human life; in fact, the commonsense reasoning problem is widely regarded as one of the most challenging in the field. There are three major problems that we must face to build a commonsense reasoning system:

Representing diverse varieties of knowledge. First, we need to find ways to *represent* in machines the kinds of commonsense knowledge that people possess. What are the kinds of data structures and vocabulary elements that are needed to represent the vast span of things that people can think about—for example, about social, economic, political, psychological, mathematical, and other types of matters? Expressing these kinds of ideas within our machines in a way that makes commonsense reasoning possible has been a challenge. Davis reviews some of the known knowledge representation techniques in (Davis, 1990), and the Cyc project (Lenat, 1995) has built a vast ontology of logical terms that can be used to describe a variety of commonsense situations, but there is still a long way to go to build effective commonsense knowledge representations for machines.

Acquiring sufficiently large knowledge bases. Second, we need to find ways to *acquire* enough commonsense knowledge about the way the world works to approximate what a typical person knows. It has been estimated that by adulthood people possess tens of millions of fragments of knowledge (Mueller, 2001), but no technique of machine learning or knowledge acquisition has been able to acquire this much knowledge. While people learn many things by living in the world, computers do not really 'live' in the world in the sense that they cannot see or manipulate things in the world nearly well as people can, and even the best algorithms for machine learning remain very weak when compared to the ability of a young child to rapidly acquire information about the world. At the same time, simply programming these pieces of knowledge is a slow and tedious process, and this 'knowledge bottleneck' is one of the major factors that has prevented AI technologies from seeing practical use.

Reasoning flexibly with commonsense knowledge. Third, we need to find ways to *reason* with commonsense knowledge, so that we can flexibly apply it to new situations. It is only in narrow, circumscribed areas that today's reasoning technologies compare to people—for example, in chess playing—where it is possible to state precisely what knowledge is needed and how to use it to perform effectively. But when it comes to more general commonsense reasoning, such as the kind needed to understand a simple children's story, it has been very difficult to write programs that can use knowledge as flexibly as people do—people can work with knowledge that is ambiguous, that has bugs of various sorts, and are amazingly good at jumping to conclusions based on partial information and revising our beliefs when presented with new information.

1.2 What is different about our approach?

We are developing three commonsense reasoning systems—ConceptNet, LifeNet, and StoryNet—that address each of these problems in new and unconventional ways:

They use natural language as an essential part of their knowledge representation. To represent knowledge in our systems, we have been using fragments of natural language as fundamental ingredients of the knowledge representation. Each of our systems uses an ontology based largely on ordinary English words, phrases, and sentences. In other words, rather than using precisely defined symbol such as #$Cat-DomesticAnimal, we simply use the word 'cat' by itself. The advantage of this approach is that our knowledge bases are especially easy for people to add to and inspect, and in addition, they are easy for application developers to interface to. People do not have to learn enormous and intricate new languages to use our systems, and instead can rely on the knowledge of English that they already possess. The challenge this approach is that unconstrained natural language is terribly vague and ambiguous when compared to computer languages. However, we have found that this is not a fatal flaw. We can always make a natural language expression more precise by adding more words—for example 'cat' can be replaced by 'house cat' or 'jungle cat', and it is often possible to use surrounding context to help disambiguate these terms. The use of English as a knowledge representation has also been demonstrated recently in the field of computational linguistics in a system that used a logical theorem prover on disambiguated WordNet glosses to enhance question-answering (Moldovan, 2003).

They take advantage of the World Wide Web and its citizens. To build sufficiently large commonsense knowledge bases, we have turned to the World Wide Web, both to its hundreds of terabytes of content and its hundreds of millions of citizens. We have built a series of knowledge acquisition interfaces that are designed for non-expert computer users from a wide range of ages and backgrounds, and our first such interface collected over 700,000 fragments of information from over 14,000 people across the web. Our interfaces are designed for uncluttered simplicity:

the activities guide the contribution and association of knowledge elements built from fragments of English. We have avoided the need for users to learn a complicated representation language and have developed interface designs that strike a balance between maximum expression for the user and beneficial knowledge acquisition. The idea is that our interfaces should be simple enough for a person to begin using almost immediately and with ease over an extended period of time. And most recently, we have begun to supplement these knowledge acquisition efforts by automatically mining the web for common sense information.

They employ alternative methods of reasoning and knowledge representation. Most recent work on commonsense reasoning has assumed that reasoning is done by logical theorem proving over knowledge expressed in logic (Davis & Morgenstern, 2004). The power of logic is that it is an extremely expressive language, comparable to natural language, yet it has a precise semantics and there are well-understood reasoning procedures for making logical inferences. But in our view, the great precision with which one must specify concepts in the logical approach is precisely the reason why there are no practical commonsense reasoning systems in the world today. In recent years there has been increasing interest in other methods of reasoning that are less sensitive to errors and ambiguities in the underlying knowledge base, such as reasoning with probabilistic models. However, there have been almost no attempts to apply these techniques to the problem of commonsense reasoning with large knowledge bases. Our commonsense reasoning systems are based not on logically sound inferencing techniques, but instead on several unsound but practically useful techniques such as spreading activation in semantic networks, probabilistic inferencing in graphical models, and case based reasoning using story-scripts. We are also working to combine these techniques into an integrated commonsense reasoning system that uses multiple representations.

In the following sections we will describe ConceptNet, LifeNet, and StoryNet in more detail. But first we will briefly describe the Open Mind Common Sense project, the predecessor to these systems and the project that launched our efforts in the area of building practical commonsense reasoning systems.

2 Open Mind Common Sense

Our efforts to build machines with common sense began in earnest four years ago with the launch of the Open Mind Common Sense web site. At the time there was only one large-scale commonsense knowledge base, the well-known Cyc knowledge base (Lenat 1995). Marvin Minsky was a great supporter of the Cyc project, but at the same time had been encouraging us and the rest of the world to start alternative projects to Cyc. He has long argued that giving computers common sense was the central problem of Artificial Intelligence, and that a problem of such importance could not be left to just one group. In any case, Cyc still had a long way to go—even though they had collected over a million units, they were predicting that they would need on the order of 100 million to engage in commonsense reasoning at the human level (Anthes, 2002).

We were interested in the question of whether it was possible to distribute the problem of building a commonsense knowledge base across thousands of people on the web, and especially, people with little or no special training computer science or artificial intelligence. We were interested in whether the ‘average person’ could participate in the process of building a commonsense knowledge base. After all, every ordinary person possesses the kind of common sense we wish to give our machines! The conditions seemed ripe to pursue such an effort. The success of distributed knowledge engineering projects like the Open Directory project was clear, and it seemed to us the only question was whether an interface could be built that the general public would find engaging enough to teach common sense, for at the time, it was not clear whether any practical applications and other benefits would ensue from such an effort.

The main question was whether there was a way for those people to express knowledge in a way that a machine could use. Unless we could find a way for people to contribute knowledge that a machine could use, then it would be a moot point if people on the web wanted to participate but found the knowledge acquisition interface too difficult to use. We began to explore the idea of using English itself as a knowledge representation language. Could people teach machines knowledge in the form of simple English statements? If so this would tremendously accelerate any effort to build a commonsense knowledge base. There would be no special knowledge representation language to learn, and so the knowledge could be contributed much more naturally. While there were clearly problems with this idea—natural languages do not have a precise formal semantics, the words in natural languages are ambiguous, and natural languages may lack words for many important common sense ideas—we decided to forge ahead since the cost of trying seemed so low. It was just a matter of putting up a web site and seeing what happened.

We built the Open Mind Common Sense (OMCS) web site to explore these ideas. OMCS was built in the first half of the year 2000, and launched in September 2000. We quickly gained an audience and as of March 2004 about 14,000 people have entered nearly 700,000 items of knowledge. The contributed knowledge consists largely of the kinds of simple English assertions shown in Figure 1, a screenshot of the OMCS knowledge browser.

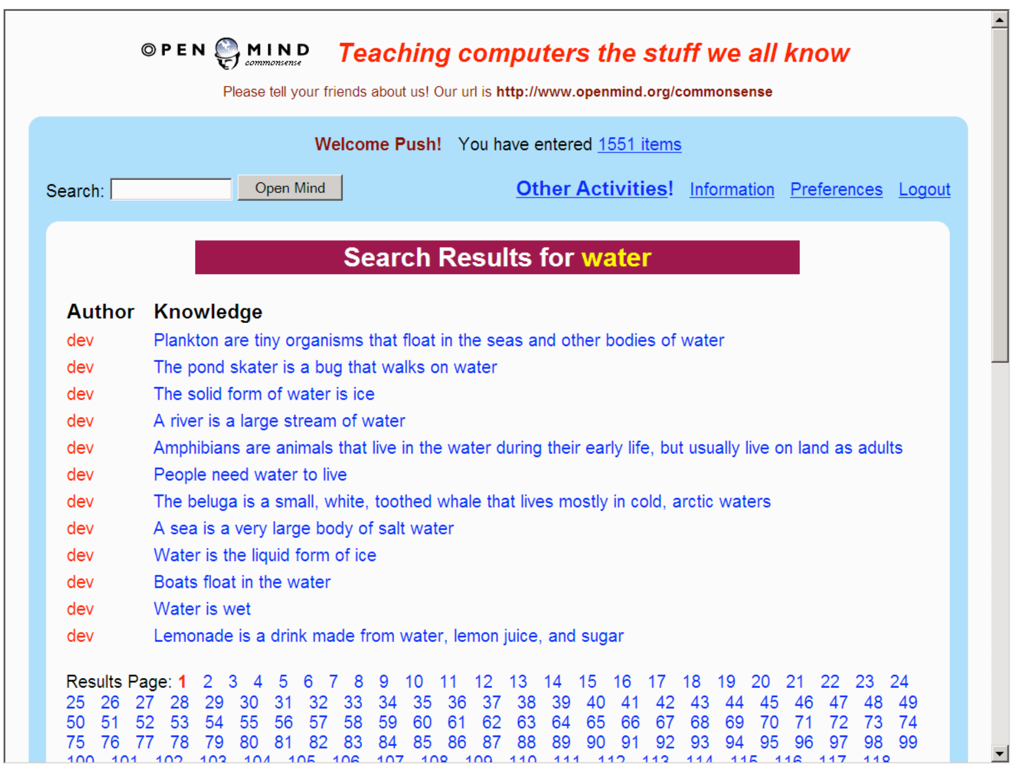

Figure 1. A few things Open Mind Common Sense knows about water

The knowledge gathered by OMCS is not a well-defined knowledge base in the sense of Cyc, which consists of clearly defined knowledge elements and an associated inference procedure.

Rather, it is a kind of English corpus—printed out, it would exceed 25,000 pages! Perhaps the best way to think about it is an encyclopedia of ordinary things, with thousands of facts about common concepts like ‘money’ (7500 facts), ‘car’ (12000 facts), ‘hair’ (3500 facts), and so forth. Many of these facts are expressed in sentences that have a simple syntax or are otherwise highly structured due to use of template-based acquisition activities.

We will not go into further detail about the content of the OMCS corpus here, but encourage the reader to visit the open mind site (at http://openmind.media.mit.edu) to browse the knowledge people have contributed. What has surprised us most of all is the high quality of this knowledge. An early analysis (Singh et al. 2002) showed that 90% of the contributed statements were rated 3 or higher (on a 5 point scale) along the dimensions of *truth* and *objectivity*, and about 85% of the statements were rated as things anyone with a high school education or more would be expected to know. Thus the data, while noisy, was not entirely overwhelmed by noise, as we had originally feared it might, and also it consisted largely of knowledge one might consider shared in our culture. Even though we were encouraged by quality ratings of the OMCS corpus, in subsequent work we developed strategies for filtering poor quality contributions using statistical methods and user assessment of contributions.

The main lesson to draw from OMCS is that the project, along with other related projects including 20Q.net, Mindpixel.com, and most recently ESPgame.org, have been successful for collecting millions of units of knowledge from members of the general public, and have proven that there is an audience for distributed AI efforts of this sort. However, there is a difference between building a large corpus of commonsense facts and building a useful commonsense knowledge base. How could we use the collected OMCS knowledge? In the next sections we describe the OMCS corpus as a resource for our commonsense reasoning systems and as a support for some of our other knowledge acquisition projects.

3 ConceptNet

The statements from OMCS that have proven the most useful are the ones that people contribute via simple templates like “flashlights can be used to ______”, or “the effect of drinking coffee is ______”. A substantial portion of the OMCS corpus consists of these kinds of semi-structured sentences, and from these we have been able to extract more usable knowledge bases.

The first knowledge base we extracted this way was ConceptNet, a large-scale semantic network mined from the OMCS corpus using a library of lexico-syntactic pattern matching rules. A fragment of ConceptNet is shown in Figure 2 below.

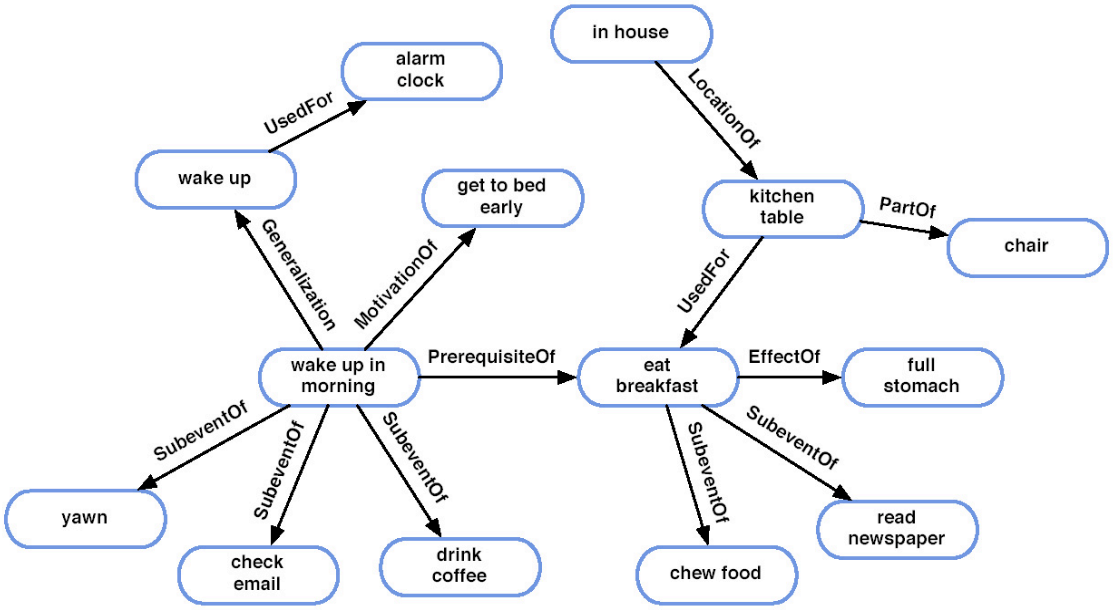

Figure 2. A subset of ConceptNet

The present version of ConceptNet consists of 1.6 million links interrelating 300,000 concepts. The links are drawn from the ontology of links shown in Table 1, and the concepts consist of simple English expressions of objects, events, properties, and places, as shown in Table 2.

Table 1. ConceptNet's Relational Ontology of 20 Link-Types

ConceptuallyRelatedTo	IsA	FirstSubeventOf	DesirousEffectOf
ThematicKLine	MadeOf	SubeventOf	UsedFor
SuperThematicKLine	DefinedAs	LastSubeventOf	LocationOf
CapableOfReceivingAction	CapableOf	PrerequisiteEventOf	MotivationOf
PropertyOf	PartOf	EffectOf	DesireOf

Table 2. Ontology of Concept Types

Events	**Things**	**Places**	**Properties**
Eat sandwich	Orange juice	At zoo	Furry
Sell car	Morning coffee	On table	Very expensive
Tell story	Policeman	Near school	Dark
Go to zoo	Leaf blower	Inside oven	Quickly
Type letter	Laptop computer	In closet	Dark

Structurally, the closest analogue to ConceptNet is the widely used WordNet semantic network (Fellbaum, 1998). ConceptNet and WordNet are both large-scale semantic networks with natural language expressions at the nodes, related by a small collection of link types. However, there are several important differences between the two systems. Whereas WordNet contains only explicit knowledge of the sort you would see in a dictionary, ConceptNet contains a much wider range of the more implicit, between-the-lines, knowledge of the sort you cannot find in any existing linguistic resource. For example, WordNet states that a cat is a type of mammal, but it says nothing about a cat being a common pet—simply because the latter statement cannot be said to be part of what 'defines' a cat. All cats are mammals, but only many cats are pets—and despite the obvious utility of the latter piece of knowledge it is excluded from WordNet on the basis that it is not always true. This points to a substantial difference between the goals of the two systems—the goal of ConceptNet is not so much to precisely define words as it is in WordNet, but rather to provide the kind of additional context and 'usually true' relations that exist between concepts.

ConceptNet and WordNet also differ in the details of their node and link types. ConceptNet includes not just words but also larger compound expressions such as "buy ticket" or "at movie theater", which lets it express more specific types of actions, objects, places, and properties. In addition, its link types go beyond WordNet's largely taxonomic ontology to include functional, temporal, spatial, affective and several other types of relations shown in Table 1. WordNet, however, has the advantage of being sense-disambiguated, and we and others are seeking to disambiguate ConceptNet as well (Chklovski & Mihalcea, 2002).

Semantically, the closest analogue to ConceptNet is the Cyc system. However, unlike Cyc, ConceptNet is very restricted in its vocabulary of predicates, and while this makes the knowledge in ConceptNet more ambiguous than the knowledge in Cyc, we have found that as a result it is very easy for people to use. Our experience with Cyc is that it takes many months to get up to speed on its ontology, and even then it is difficult to keep track of how to express complex ideas—it is very much like learning a new natural language with a comparably sized vocabulary. Perhaps the best example of the ease of use of ConceptNet was a course that Henry Lieberman taught at the Media Lab, where the students for their final projects built a variety of systems using an early version of ConceptNet. These students largely had little background in AI besides the introductory course, but nonetheless managed to build a variety of interesting systems in only a matter of weeks.

We reason with ConceptNet using various methods of traversing links or 'spreading activation' to retrieve related concepts. For example, given a concept like 'drink coffee', we can retrieve related concepts like 'pour coffee', 'make coffee', 'feel awake', and so forth. So given a set of starting concepts, we can begin to traverse the network to find related fragments of information. Using such methods we can:

- infer what events might come next
- infer what might have happened earlier
- infer what objects might be required to perform an action
- infer what the properties of objects are
- infer where an object might be found
- infer what sorts of goals people might have

More information about ConceptNet and the kind of reasoning it supports is given in (Liu & Singh, 2004), a companion article in this volume. Generally, this sort of reasoning can be used to elaborate search queries, recognize situations and events from their elements, find the semantic similarity between passages of text, and many other applications. We have built a variety of applications using ConceptNet such as a photo retrieval agent, an affective text classifier and a topic spotter for conversations. Details of these and other applications that use ConceptNet can be found in (Lieberman et al., 2004).

4 LifeNet

While ConceptNet was a very successful system, as measured by the many applications that used it, two problems remained difficult to solve. First, the meanings of its nodes and links are ambiguous, and it is unclear how to take that ambiguity into account during reasoning. Second, there are a number of erroneous links, and it seems highly unlikely that we will ever be able to get rid of all the errors if we are collecting knowledge from a wide range of users. Is there a method of reasoning that can tolerate such ambiguities and errors?

We began to explore probabilistic methods for representing and reasoning with commonsense knowledge. Probabilistic methods for reasoning are more tolerant than rule-based reasoning methods to the uncertainty in our knowledge of the situation, as well as to the uncertainty in the reliability of the rules themselves. Also there are simple and well-understood inference procedures that can run fairly fast if you are willing to accept approximate solutions.

However, probabilistic approaches have a major disadvantage when it comes to commonsense reasoning. To date, there are only a handful of poorly understood techniques for reasoning with the kinds of expressive representations that are needed to express a wide range of commonsense situations and events, for example, using frames, scripts, or first-order logic. Is it possible to build a commonsense knowledge base using less expressive, propositional representations?

We developed the LifeNet system to explore these issues. LifeNet is a large-scale temporal graphical model expressed in terms of egocentric propositions of the form:

- I am at a restaurant
- I eat a sandwich
- It is 3 pm
- It is raining outside
- I feel frightened
- I am drinking coffee

Each of these propositions is a statement that a person could say was true or not true of their situation, perhaps with some probability. In LifeNet these propositions are arranged into two columns representing the state at two consecutive moments in time, and these propositions are linked by joint probability tables representing both the probability that one proposition follows another in time, and also the probability of two propositions being true at the same time. A small sample of LifeNet is shown in Figure 3 below.

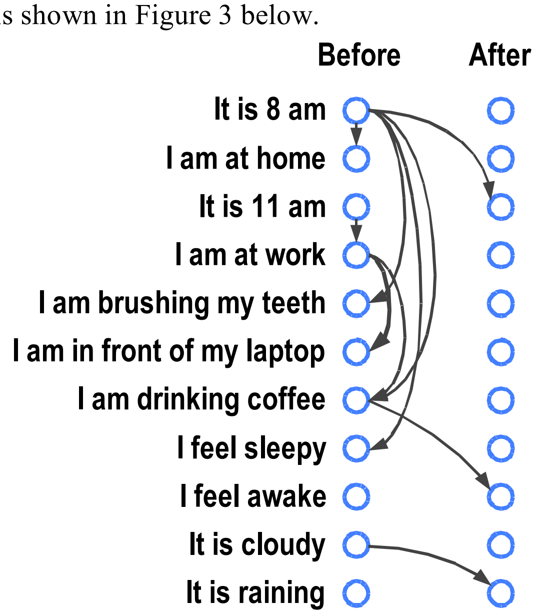

Figure 3. A sample of LifeNet. The before column shows t1 and the after column shows t2. "It is 8am" occurs before "It is 11am". "It is 8am" occurs at the same time as "I am brushing my teeth."

LifeNet can be regarded as an approximate, first-person model of the course of everyday human experience. Presently it consists of a total of 80,000 propositional nodes linked by 415,000 joint probability tables between pairs of nodes. These nodes and linked are generated by transforming a subset of ConceptNet, a process described in more detail in (Singh and Williams, 2003). Unlike ConceptNet, the semantics of the nodes and links in LifeNet are quite clear. The links represent probabilistic constraints, both temporal and atemporal, between propositional fluents each representing one measurable aspect of the world. LifeNet is a *Markov network*, an undirected graphical model where the nodes represent random variables and the edges joint probability constraints relating those variables. We reason with LifeNet using local belief updating techniques, in particular, 'loopy' belief propagation as described by Pearl (Pearl, 1988). Using these techniques, we can use LifeNet to engage in several types of temporal reasoning:

- **Prediction**: Guess what might be true in the next moment.
- **Elaboration**: Guess what else might be true now.
- **Explanation**: Guess at what happened prior to the current event.
- **Projection**: Guess what series of events might follow.
- **Filtering**: Filter unlikely current states or events.
- **Fixed-lag smoothing**: Filter unlikely past states or events.

LifeNet is less expressive than ConceptNet in the sense that it focuses exclusively on the details of situations and transitions between situations. But in exchange for this loss in expressivity, LifeNet provides a way to tolerate errors in the knowledge base, and provides a knowledge representation with a clear semantics and inference procedure for reasoning with the collected knowledge.

We are using LifeNet to help build more context-aware systems. Traditionally there has been much emphasis on building systems that can *sense* various features of the world, ranging from faces to sounds to emotions, and react to these sensations in useful ways. But all such projects have been limited in the quality of the systems' responses because *interpretation* requires more common sense. Bridging common sense and sensing requires finding ways to connect the typically probabilistic representations used by perceptual systems and the typically non-probabilistic representations used by commonsense reasoning systems. Because LifeNet is based on probabilistic representations, it is straightforward to connect to lower level sensory systems. Eagle (Eagle et al., 2004) describes an initial foray into this area of enhancing sensory systems with common sense, and we hope to do more work in this area.

4.1 LifeNet Interface

LifeNet was built using knowledge from ConceptNet, but as with OMCS we have built an easy-to-use web site that allows an untrained person to quickly add more knowledge to LifeNet. The interface is shown in Figure 4. At present the interface allows the user to supply knowledge about both temporal and atemporal links by dragging and dropping propositions into boxes denoting truth and falsity over a series of time steps.

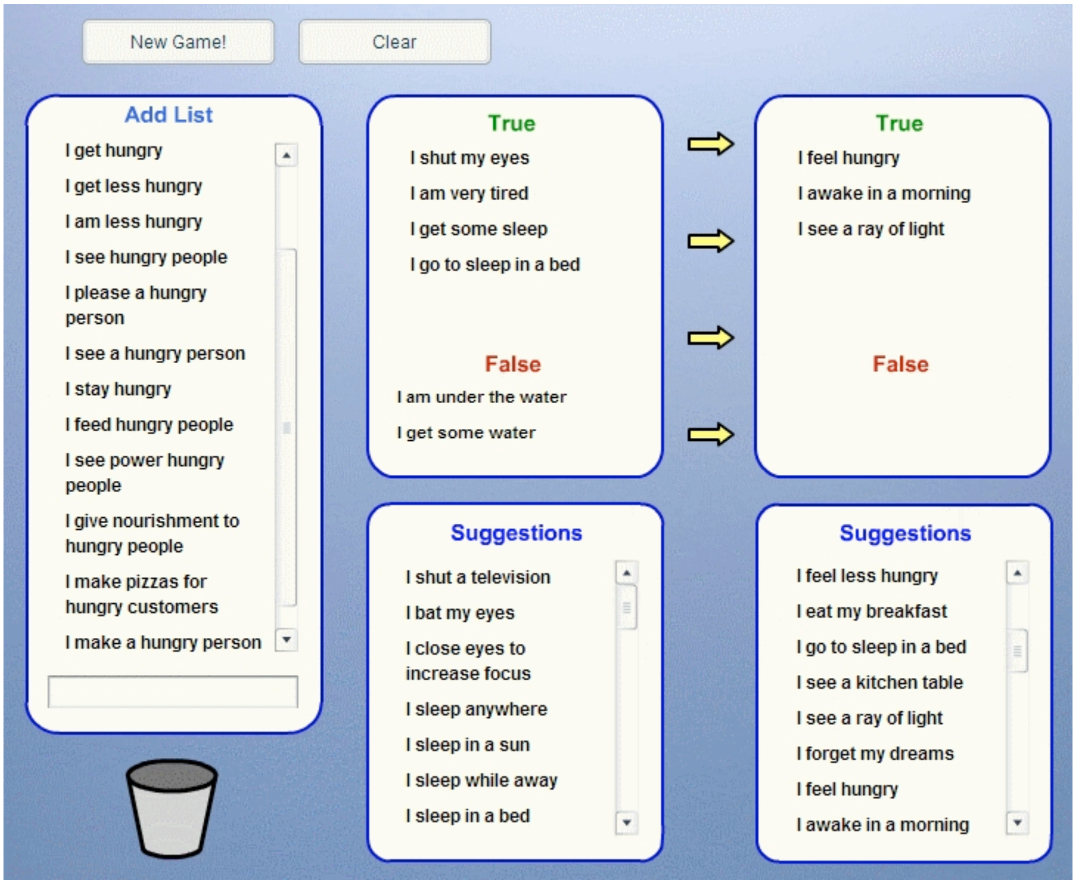

Figure 4. Interacting with the LifeNet inference engine

The interface is divided into three areas. The "Add List" allows the users to directly search for propositions in the knowledge base. The "True/False" boxes are the composition areas that delineate a series of time steps. Users can add or delete propositions when building a situation for submission. The third area is the "Suggestion" area, which displays plausible propositions for each time step generated by the LifeNet inference engine. Essentially the system is guessing what else might be true at the same time as the selected propositions. Suggestions can then be added to any "True/False" box or ignored.

The goal of the user is to complete a game by building a situation comprised of true and false propositions. Games can be initiated by the system by display of a few starting propositions or the user can clear the board and begin a game by searching the add list for any proposition or concept they like.

We discovered the major challenges in building such an interface are supplying reasonable suggestions, providing flexibility in composition and promoting long-term engagement. Many of the generated suggestions are reasonable new candidates because the existing inferences are only partly reliable. As the system is used we expect the suggestions to become even more robust. We are addressing the need for more compositional flexibility, as requested by users, by adding the ability to extend the number of time steps and contribute new propositions, guided by templates. There is great potential for long-term participation since LifeNet engages the natural human ability to recall and order our life experiences.

The advantage of LifeNet is that its representation scheme is very simple and we believe that such systems will prove very useful in building richer models of human behavior. However, ultimately the impoverished nature of this representation makes it difficult to capture higher level and longer term ideas about ordinary human life, and especially, about the complex types of relationships that occur in stories.

5 Representing common sense using stories

ConceptNet and LifeNet represent common sense in small units of knowledge linked by binary relations and joint probability tables. However, using such small units of knowledge forces one to ignore many complex and subtle aspects of how the real world works

What knowledge representation can we use to approach the complexity of the real world? Stories are a type of representation people use to organize their lives and make sense of the world. A story can be thought of as pattern within which events or series of events can be ordered, understood and communicated (Livo & Rietz, 1986). Richard Kearney reminds us of the long-standing recognition in philosophy of the cycle of action and narrative (Kearney, 2002). As we live we create life-stories in order to remember our past and project our future, a process first named "poesis" by Aristotle. The creation of life-stories involves the structural plotting of information (Aristotle, 1968). Aristotle regarded a successful plot as one in which the episodes followed each other in probable and necessary sequence. The relations between elements are as important as the elements themselves. Here we take our cue for creating complex commonsense knowledge representations. While it is the job of poets and artists to reinvent and test the limits of plotting life, it is our job to find a story representation that will support machine understanding of everyday life. We define a story as a set of representational elements and their relations. Representational elements include actors, events, states, settings and objects. Relationships between elements can express goals of the actors, problems to be solved, plans that are successful or unsuccessful and, at their most complex, themes.

As a representation stories have other advantages besides their capacity to depict complexity. Stories are 'implicit contexts' for knowledge, which have many of the advantages of solely explicit contexts. Consider this example scenario. Jack cannot find his keys. A story about the situation might look like this:

- It was 11 pm (when would you need a flashlight?)
- Jack couldn't find his car keys (what problem might a flashlight help with?)
- Jack looked in the yard (where would you use a flashlight?)
- It was too dark to see anything (why would you need a flashlight)
- He got a flashlight from a drawer (where would you find a flashlight?)
- Now he could see the yard (what does a flashlight do?)
- He found his car keys in the grass (problem solution)

While ConceptNet can be used to generate knowledge such as 'keys are used for opening something' and 'yards are located near houses' and LifeNet can offer that 'in a yard' often occurs in the same situation as 'see the sky', neither can directly provide inference to help solve the problem of lost keys. Stories supply more detail, keep related knowledge together, and by connecting knowledge about objects, events, places, functions, goals, etc., they can help direct inference. Another strength of stories is the capacity for generalization. Using case-based reasoning a story about finding keys can be generalized to a story about finding.

Just as semantic networks and probabilistic graphical models have been little applied to large-scale commonsense reasoning problems, so has the use of story-like structures. While in the early days of AI there was much interest in expressing commonsense knowledge in terms of large, story-like units (Minsky, 1974; Schank and Abelson, 1977), recently such approaches have rarely been used for commonsense reasoning. Mueller is a notable exception (Mueller, (in press)). No present large-scale semantic knowledge bases contain a substantial amount of story knowledge. Mueller compared several systems (Cyc (Lenat, 1995), FrameNet (Johnson, 1998), Gordon's Expectation Packages (Gordon, 1999), ThoughtTreasure (Mueller, 1998), and WordNet 1.6 (Fellbaum, 1998) and found that these systems consisted largely of facts and rules, and not cases and stories against which case-based reasoning could be performed (Mueller, 1999). A large-scale story knowledge base would be a fundamentally new kind of resource.

In addition, from the perspective of turning to the general public to build knowledge bases, we suspect that the average person may be better at telling and explaining simple stories than generating more direct forms of knowledge engineering such as formulating and encoding logical rules. Since people are by nature experts at using story structures to organize and manage the complexity of life, it may be easier to tell and explain a specific story, which focuses the user on a specific set of characters, objects, and events, and their relationships, than to make a general rule-based theory in the abstract of some domain.

In the next sections we describe two of our first approaches to the problem of building a large semi-structured story knowledge base. As our acquisition systems are released and tested we anticipate that our representations will be revised and enhanced as we learn from the human storytelling impulse and the success and limitations of the inferences mechanisms we develop.

5.1 Open Mind Experiences

During the development of Open Mind Commonsense we had the notion that useful stories could perhaps be generated automatically from the temporally-ordered and causally-connected knowledge in the OMCS corpus. We built a system called MAKEBELIEVE (Liu & Singh, 2002) that generated simple first-person storylines collaboratively with a person by alternating turns. The stories generated by MAKEBELIEVE were often quirky and silly because while the system had a notion of causal-connectedness, it did not know how to assemble story pieces to satisfy a global story context, which requires notions like relevance, and hiding versus highlighting. People are much better at building coherent and meaningful stories.

ConceptNet and LifeNet both approached the idea of a story-like representation for commonsense reasoning. ConceptNet includes the temporal relations such as 'the-first-thing-you-do' and 'the-last-thing-you-do' that suggest the beginning and end of a story. LifeNet can predict simultaneous, preceding or following events in a situation. However, this knowledge is fairly sparse in ConceptNet and LifeNet, and so we decided to try to build a web site that explicitly collected knowledge in the form of stories.

Open Mind Experiences (OMEX) was a first attempt to gather structured story knowledge from the general public (Singh and Barry, 2003). In many ways, OMEX improved upon and fixed several problems with the original OMCS web site OMEX focused exclusively on knowledge entry through template activities, so all the collected knowledge was guaranteed to be at the very least semi-structured. The story contribution templates were hand-built and are a thematic representation based on Lehnert's plot units (Lehnert, 1982). Plot units are a convenient way to represent a wide range of story types as linked positive events, negative events and mental states. Lehnert used this representation as a way of identifying central concepts of a story plot during text summarization. We created templates based on plot units to prompt acquisition of multi-character stories across broad subject domains. Plot unit representations also have the advantage of being compositional. Simple plot units can be combined to create even more complex story representations. The development of OMEX was especially valuable because it addressed the need for social commonsense. Contributors to OMEX entered knowledge into simple (one character) and complex (multi-character) entry templates. The design prompted users to enter stories that expressed themes such as competition and cooperation.

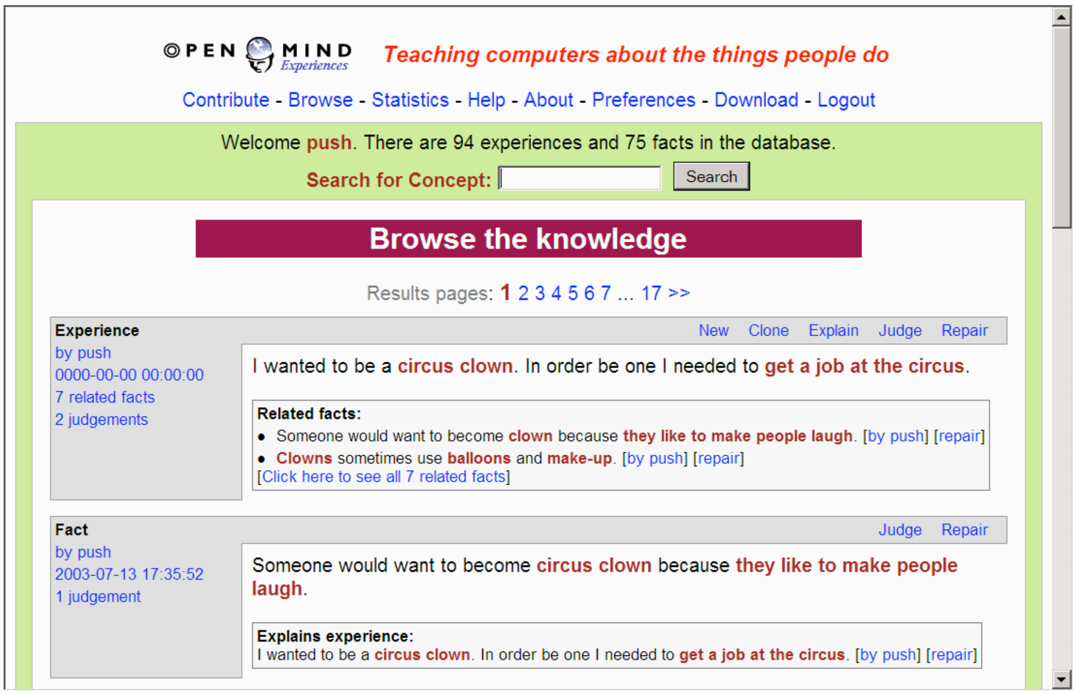

Figure 5. Browsing the Open Mind Experiences web site

OMEX also introduced the idea of explanation, judgment and repair of knowledge by contributors, so that erroneous or controversial facts could be marked as dubious, corrected, contextualized, or further commented upon. Users could explain a story by answering some questions that solicit the background knowledge needed to understand and experience. Users could also classify stories along various dimensions, such as plausibility or typicality, and correct grammatical errors.

The goal of the OMEX site was essentially to collect two general classes of story information, the story content and the story meta information.

5.1.1 Story content

Studies of human and computer memory and understanding have yielded many different story content theories, from simple sets of events to very complex configurations. There are a variety of concepts that must be addressed in designing a story representation. One must make a commitment to a set of symbols and a method for combining them to enable robust inference. This is particularly difficult if one does not have a target story in mind that the resulting system will need to understand. Since our target is to understand the broadest range of stories we looked to previous work in story understanding as the basis for our representations. Mueller provides a helpful general overview of symbolic knowledge structures for story understanding that we recap here (Mueller, 2002):

- **Script:** Temporal list of causally linked events.
- **Plan:** Steps executed to achieve a goal.
- **Goals:** Specifications of a desired state.
- **Theme:** A pattern of linked goals, states and events.
- **Grid:** Locations of objects and characters in two-dimensions.
- **Time:** Absolute or relative descriptions of time.

Currently we focus on gathering content in the form of simple scripts, goals, themes and relative descriptions of time. As we add to the activities in our acquisition sites we are addressing the remaining historical structures and developing other new representations to aid commonsense reasoning.

5.1.2 Story meta information

While story content expresses what happens in a story, meta information addresses some subtleties of how a story is used. In (Lenat, 1998) Lenat argues that it is important to encode not just facts and rules about domains, but also meta-assertions that describe precisely in what situations those facts and rules apply, what other knowledge may be relevant, what problem-types those facts and rules may help solve, and so forth. Consider this list of story attributes to direct commonsense reasoning.

- **Author information**:
  Age, ethnicity, class, gender, occupation and residence
- **Expectation information**:
  Does it happen every day?
  Is this a story that everyone has experienced at some time or another?
  Is this an unusual story?
- **Story topic**:
  What is the story about?
  What are the most important terms, goal expressions and themes?
- **Story conditions**:
  What are the prerequisites of this story?
  What conditions will initiate this story?
  What conditions in a situation would make this story impossible?
- **Story uses**:
  Does this story teach you how to solve a problem?
  Does it tell you to avoid a situation?
  Does it present a model of a person who you could emulate?
- **Story realms**:
  What realms of commonsense are expressed in this story?
  Does it have information about the physical world, the social world or both?

OMEX had an activity where people could teach some of this knowledge, but we are still looking into good ways to represent and collect this type of knowledge. We also anticipate that the collection of this information will not only help our commonsense reasoning efforts but also provide a resource for psychologists who are studying how people make inferences in order to comprehend texts, films and the everyday world (Graesser et al., 1994).

5.2 StoryNet

While OMEX was in many ways a substantial improvement over the original OMCS site, launching a test version of the OMEX site revealed that the knowledge entry templates were too abstract to engage the typical user and lacked the constraints necessary to support successful knowledge extraction. Ultimately, these shortcomings convinced us to the design a new type of interface based on suggestions, sorting and constrained text entry. Recognizing these problems helped us rethink the OMEX interface, and ultimately led us to the StoryNet system, which is currently under development.

Just as LifeNet built on ConceptNet, StoryNet builds on LifeNet and ConceptNet, using them as resources. While ConceptNet lays out the possibilities for ordering elements, not all combinations will make sense. If you start at a random node in ConceptNet and follow it to generate a story, you might get something like this:

> *I want to drive a car.*
> *I need gasoline.*
> *Gasoline can be found in a plane.*
> *A plane can be found in the sky.*

Links between concepts are inherently local in time and do not factor in more constraints or are organized into larger units. The chaining of ConceptNet events is not directed by the goal of driving a car. In very few steps the inference goes off on a tangent that, however amusing, unnecessarily and nonsensically complicates the action.

StoryNet uses the same ontology of concepts as LifeNet but captures knowledge in the form of scripts, lists of temporally ordered events. Generally, our work with both OMEX and StoryNet inherits strongly from the early work in story understanding by Schank and his colleagues (Schank and Abelson, 1977). Schank introduced scripts as memory structures for reasoning about story texts. In Schank's representation there was a canonical script for each activity, for example the classic restaurant script. StoryNet scripts are not intended to be general; rather they are instances, general descriptions of activities. Thus, there can be many 'restaurant scripts' in the StoryNet knowledge base. The scripts in StoryNet contain only sequences and we rely on our other commonsense resources, ConceptNet and LifeNet, to generate other information of the type found in Schankian scripts - elements such as actors, emotions, objects and the conditions necessary for the script to be instantiated.

Consider this simple example:

> *I am hungry.*
> *I want to go to a restaurant.*
> *I see a car.*
> *I get into a car.*
> *I start the car.*
> *I drive to a restaurant.*
> *I enter the restaurant.*
> *I order a sandwich.*

StoryNet can offer multiple scenarios of people traveling to a restaurant – riding a bicycle, walking, running, taking a bus, etc. The improved representation allows us to tell a story—a set of descriptions, relationships, and events concerned with a specific set of characters, objects, places, etc. Event relationships are very complex. Events might be dependent or independent. 'Getting into a car' is necessary to 'drive a car.' 'Turn on the car radio' is not necessary. In some circumstances one event might subsume another or thwart the goal of a person in the story. We use the script representation as a basis for collecting more relational knowledge of this kind.

Each event in this script can be expanded to include details of the situation. In order to create a sensible story, people know that driving a car involves traveling a distance. When the driver arrives at the restaurant, she is at a different place then when she entered the car. Other details might be necessary in different contexts. If the point of the story is safety while driving, then we might need to answer questions like: 'What was the weather?' 'How much gas is the car?' 'Does the driver have good vision?' A goal of StoryNet is to create a representation that is flexible. ConceptNet and LifeNet can be used to address some of these questions by expanding the context, inserting additional steps or determining the likelihood of event-order in a story. LifeNet could provide the likelihood of a person driving to a restaurant after starting a car. ConceptNet could provide a list of objects that might be found in a restaurant.

5.2.1 StoryNet acquisition

The StoryNet interface is designed to be as harmonious as possible with human storytelling activity so that the average person can use it. StoryNet is also designed to be general-purpose so that people could use it to enter both personal information, as well as information that is more obvious, common, and shared between people. This can ultimately be useful when StoryNet is used in applications that require both general information about the world and specific information about users.

How do people interact with the system to teach it new knowledge? We are looking into several methods of acquisition. The first method of acquisition is to have people who register with the web site to describe their own life experiences. This approach could be thought of as a simple diary of everyday events. Information about the contributor is kept as a way to further evaluate these stories with respect to authorship. We can analyze the contributions and search for similarities and differences. Do people under the age of 18 submit stories of a different quality than people over the age of 60? Who tells more problem-solving stories about objects? Engineers or social workers? The background contributor helps to determining the context of the stories that are being told. Ultimately, one person's life experiences can be used to understand another's. This could be helpful in disciplines which rely on stories for effective communication such as education, narrative psychology, politics, law, and entertainment.

We have built an initial StoryNet interface based upon the LifeNet interface model. It is a simple drag and drop interface for sorting and ordering propositions. A user can search for a statement to begin a story or be prompted by the system. As each proposition is added to the story construction box, suggestions of possible next events are provided.

Figure 6. Assembling a simple story using StoryNet

Currently, we are developing constrained methods for entering new propositions that adhere to our model of four proposition types: actions, object, locations and properties. This will be accomplished by a pull down menu in each proposition object that will provide alternative suggestions and ability for user entry of test. For example, the proposition "I see a cloud" might be altered to "I see a bird."

Note that this is the first version of the interface built with the primary activity of script gathering. We are developing other activities that allow the user to interact with the story knowledge base by elaborating on a story, answering questions about it or altering it by substituting one element. We are also creating a version that puts users in competition in a story creation game. We anticipate this will be entertaining and might also improve user commitment to the system.

We are presently working on incorporating case-based reasoning into the StoryNet web site so that users can tell stories interactively, merge two stories together, chain stories to generate new stories, substitute elements for more sensible or interesting elements, and other manipulations of story structures.

6 Integrating these systems

The approaches taken by ConceptNet, LifeNet, and StoryNet each have their own focus. One way to conceptualize their interrelations is as follows:

- StoryNet scripts give you a longer-term model of the flow of events, and relate elements like the situation, goals, and other ingredients as they are connected in a particular sequence of events.

- LifeNet drills down into the details and captures much more detail about individual situations and transitions between pairs of situations than we would find in StoryNet.

- ConceptNet abstracts away from any given situation or story and instead expresses the broad collection of related concepts that would appear in a host of related stories.

These three systems are thus quite complementary. ConceptNet can express a wide range of relationships between objects and events, but lacks the ability to tie such elements together into a larger-scale situation as StoryNet does. StoryNet, however, requires that one express a sequence of events and inhibits detailed descriptions of the situation in which the sequence occurs. Those details, however, are exactly what LifeNet expresses.

Can we build a reasoning system that combines the advantages of these three commonsense reasoning systems? What is the relationship between these systems? We are beginning work on a system that combines their different reasoning techniques. Figure 7 below shows how these three different representations might be related in a unified system. We are working on a visual interface that lets us view and switch between each of these representations, in order to build an integrated browser for the three knowledge bases.

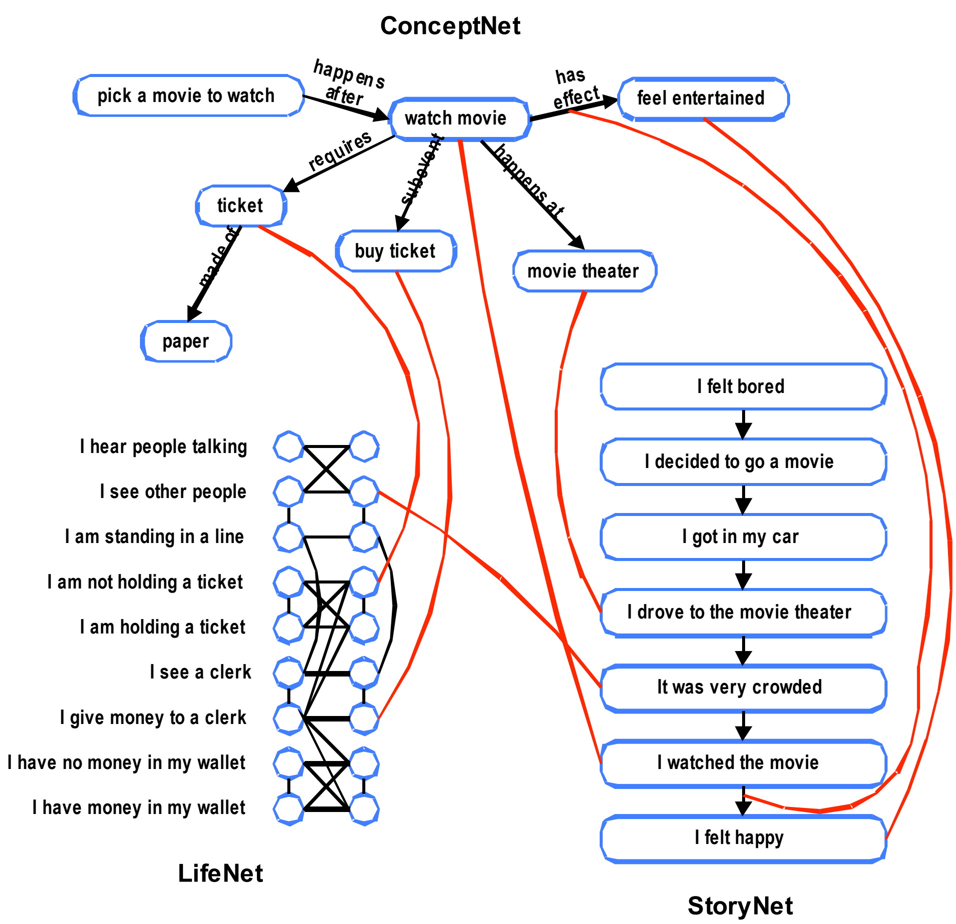

Figure 7. An example of relating concepts in ConceptNet, LifeNet, and StoryNet. Links and nodes of the different networks are connected.

7 Future Work

To get an update on current commonsense work at the MIT Media Lab, please visit our Commonsense Computing Website (http://csc.media.mit.edu).

8 Conclusions

There has been little work on the practical commonsense reasoning problem in over a generation: many in the AI community still consider the commonsense problem too difficult. At the Media Lab we have been taking a fresh look at the problem of building practical commonsense reasoning systems using unconventional techniques – representing knowledge in natural language, distributing knowledge acquisition to non-experts via the World Wide Web, and developing reasoning techniques that work successfully with large and imperfect knowledge bases. We believe the progress we have made in the last few years is a sign that the commonsense reasoning problem—long regarded as one of the most fundamental and challenging problems in the field—may be tractable after all, and that commonsense-enabled systems may become, far sooner than expected, indispensable elements of our technological society.

Acknowledgements

We would like to extend our thanks all the people who have been involved with the Media Lab common sense efforts since the beginning, and especially, to the many thousands of people who have contributed their knowledge to our databases. This work is supported by the sponsors of the MIT Media Lab.

References

Anthes, G. (2002). *Firm aims to 'computerize' commonsense*. http://www.cnn.com/2002/TECH/industry/04/11/memome.project.idg/index.html. Accessed March 28, 2004.

Aristotle, (1968). Poetics, translated by Leon Golden, Englewood-Cliffs: Prentice-Hill

Chklovski, T. and Mihalcea, R. (2002). Building a Sense Tagged Corpus with Open Mind Word Expert. In *Proceedings of the Workshop on "Word Sense Disambiguation: Recent Successes and Future Directions", ACL 2002.* pp. 116-122.

Davis, E. (1990). *Representations of Commonsense Knowledge*. San Mateo, CA: Morgan Kaufmann.

Davis, E. and Morgenstern, L. (2004). Introduction: Progress in formal commonsense reasoning. *AI Journal: Special Issue on Logical Formalizations and Commonsense Reasoning*. *153*(1-2):1-12.

Eagle, N., Pentland, A., and Singh, P. (2003). Common Sense Conversations: Understanding Conversation using a Common Sense Data-base. To appear in *Proceedings of "Artificial Intelligence, Information Access, and Mobile Computing Workshop", IJCAI 2003*. Acapulco, Mexico. August 2003.

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Gordon, A. S. (1999). *The design of knowledge-rich browsing interfaces for retrieval in digital libraries*. Doctoral dissertation, Northwestern University, Evanston, IL.

Graesser, A., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371-395.

Johnson, C. (1998). *Syntactic and semantic principles of FrameNet annotation*. University of California, Berkeley.

Kearney, Richard. (2002). *On Stories.* New York: Routledge.

Lehnert, W. (1982). Plot Units: A Narrative Summarization Strategy. In Lehnert, W., and Ringle, M. (Eds.), *Strategies for natural language processing*. (pp. 375-414). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lenat, D. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 33-38.

Lenat, D. (1998). The dimensions of context-space. Retrieved from http://www.cyc.com/context-space.doc.

Lieberman, H. and Liu. H. (2002). Adaptive Linking between Text and Photos Using Common Sense Reasoning. In *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*, (AH2002) Malaga, Spain.

Lieberman, H., Liu, H., Singh, P. and Barry, B. (2004). Beating some common sense into interactive applications. To appear in *AI Magazine.*

Liu, H. and Singh, P. (2002). MAKEBELIEVE: Using Commonsense to Generate Stories. *Proceedings of the Eighteenth National Conference on Artificial Intelligence*. AAAI Press, pp. 957-958.

Liu, H. & Singh, P. (2004). ConceptNet: A practical commonsense reasoning toolkit. BT Journal article.

Livo, N. & Ritz, S. (1986). Storytelling: Process and Practice. Englewood, Colorado: Libraries Unlimited.

Minsky, M. (1974). *A framework for representing knowledge* (AI Laboratory Memo 306). Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Mueller, E. T. (1998). *Natural language processing with ThoughtTreasure*. New York: Signiform.

Mueller, E. T. (1999). *A database and lexicon of scripts for ThoughtTreasure*. CogPrints ID cog00000555 http://cogprints.soton.ac.uk/

Mueller, E. T. (2001). *Common sense in humans*. Available: http://www.signiform.com/erik/pubs/cshumans.htm

Mueller, Erik T. (2002). Story understanding. In Lynn Nadel (Ed.), *Encyclopedia of Cognitive Science*. London: Nature Publishing Group.

Mueller, Erik T. (in press). Understanding script-based stories using commonsense reasoning. *Cognitive Systems Research*.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: networks of plausible inference. San Mateo, CA: Morgan Kaufman.

Schank, Roger C., and Abelson, Robert P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Singh, P. (2002). The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*. Palo Alto, CA: AAAI.

Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., & Zhu, W. L. (2002). Open Mind Common Sense: Knowledge acquisition from the general public. *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*. *Lecture Notes in Computer Science* (Volume 2519). Heidelberg: Springer-Verlag.

Singh, P. and Barry, B. (2003). Collecting commonsense experiences. *Proceedings of the Second International Conference on Knowledge Capture (K-CAP 2003).* Sanibel Island, FL.

Singh, P. and Williams, W. (2003). LifeNet: a prepositional model of ordinary human activity. *Proceedings of the Workshop on Distributed and Collaborative Knowledge Capture* (DC-KCAP) at K-CAP 2003.