ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge

Catherine Havasi
Laboratory for Linguistics and Computation
Brandeis University
415 South Street
Waltham, MA 02454
havasi@cs.brandeis.edu

Robert Speer
Common Sense Computing Group
MIT Media Lab
20 Ames Street
Cambridge, MA 02139
rspeer@mit.edu

Jason B. Alonso
Tangible Media Group
MIT Media Lab
20 Ames Street
Cambridge, MA 02139
jalonso@media.mit.edu

Abstract

The Open Mind Common Sense project has been collecting common-sense knowledge from volunteers on the Internet since 2000. This knowledge is represented in a machine-interpretable semantic network called ConceptNet.

We present ConceptNet 3, which improves the acquisition of new knowledge in ConceptNet and facilitates turning edges of the network back into natural language. We show how its modular design helps it adapt to different data sets and languages. Finally, we evaluate the content of ConceptNet 3, showing that the information it contains is comparable with WordNet and the Brandeis Semantic Ontology.

Keywords

Knowledge representation, common-sense reasoning, natural language processing, information extraction

1 Introduction

Understanding language in any form requires understanding connections among words, concepts, phrases and thoughts. Many of the problems we face today in artificial intelligence depend in some way on understanding this network of relationships which represent the facts that each of us knows about the world. Researchers have looked for ways to automatically discover such relationships, but automatic methods can miss many basic relationships that are rarely stated directly in corpora. When people communicate with each other, their conversation relies on many basic, unspoken assumptions, and they often learn the basis behind these assumptions long before they can write at all, much less write the text found in corpora.

Grice's theory of pragmatics [5] states that when communicating, people tend not to provide information which is obvious or extraneous. If someone says "I bought groceries", he is unlikely to add that he used money to do so, unless the context made this fact surprising or in question. This means that it is difficult to automatically extract common-sense statements from text, and the results tend to be unreliable and need to be checked by a human. In fact, large portions of current lexical resources, such as WordNet, FrameNet, PropBank, Cyc, SIMPLE and the BSO, are not collected automatically, but are created by trained knowledge engineers. This sort of resource creation is labor intensive and time consuming.

In 2000, the Open Mind Common Sense project began to collect statements from untrained volunteers on the Internet. Since then, it has amassed over 700,000 pieces of information both from free and structured text entry. This data has been used to automatically build a semantic network of over 150,000 nodes, called ConceptNet. In this paper we introduce ConceptNet 3, its newest version. We then compare information in ConceptNet to two primarily hand-created lexical resources: the Generative Lexicon-inspired Brandeis Semantic Ontology project [13] and WordNet [4].

2 The Open Mind Common Sense Project

The Open Mind Common Sense (OMCS) project serves as a distributed solution to the problem of common sense acquisition, by enabling the general public to enter common sense into the system with no special training or knowledge of computer science. The project currently has 14,000 registered English language contributors.

OMCS collects data by interacting with its contributors in activities which elicit different types of common sense knowledge. Some of the data is entered free-form, and some was collected using semi-structured frames where contributors were given sentences and would fill in a word or phrase that completed the sentence. For example, given the frame "____ can be used to ____.", one could fill in "a pen" and "write", or more complex phrases such as "take the dog for a walk" and "get exercise".

Open Mind Commons [15] is a new interface for collecting knowledge from volunteers, built on top of ConceptNet 3, which allows its contributors to participate in the process of refining knowledge. Contributors can see the statements that have previously been entered on a given topic, and give them ratings to indicate whether they are helpful, correct knowledge or not. Also, Commons uses the existing knowledge on a topic to ask relevant questions. These questions help the system fill in gaps in its knowledge, and also help to show users what the system is learning from the knowledge they enter.

Each interface to OMCS presents knowledge to its users in natural language, and collects new knowledge in natural language as well. In order to use this knowledge computationally, it has to be transformed into a more structured representation.

2.1 The Birth of ConceptNet

ConceptNet is a representation of the Open Mind Common Sense corpus that is easy for a variety of applications to process. From the semi-structured English sentences in OMCS, we are able to extract knowledge into more computable representations. When the OMCS project began using the data set to improve intelligent user interfaces, we began employing extraction rules to mine the knowledge into a semantic network. The evolution of this process has brought us to ConceptNet 3.

In this version of ConceptNet, we focus on the usefulness of the data in the OMCS project to natural language processing and artificial intelligence as a whole. We have aimed to make ConceptNet modular in a way which enables us to quickly and easily make ConceptNets for other data sets such as the Brazilian Open Mind. To support this change of focus, improvements such as higher-order predicates, polarity and improved weighting metrics have been introduced.

ConceptNet and OMCS are useful in a wide variety of applications where undisambiguated text is used. One example of this is improving the accuracy of speech recognition [8]. ConceptNet can also be used to help an intelligent user interface understand the user's goals and views of the world [9]. For use of ConceptNet 3 as an evaluative tool please see [6]. An extensive summary of applications using the ConceptNet framework can be found in [10].

2.2 Multilingual Knowledge Collection

In 2005, a sister project to Open Mind Common Sense was established at the Universidade Federal de São Carlos, in order to collect common sense knowledge in Portuguese [2]. The *Open Mind Commonsense no Brasil* project has now collected over 160,000 statements from its contributors. GlobalMind [3], a project to collect similar knowledge in Korean, Japanese, and Chinese and to encourage users to translate knowledge among these languages and English, was launched in 2006. These projects expand the population that can contribute to Open Mind, and give us the potential to build connections between the knowledge bases of the different languages and study the cultural differences that emerge.

ConceptNet 3 is flexible enough with its natural language tools that it can build ConceptNets for multiple languages and synthesize them into the same database. We have now done so with the Portuguese corpus, which is the most mature of OMCS' sister projects.

2.3 OMCS and Other Resources

2.3.1 Cyc

The Cyc project [7] is another attempt to collect common sense knowledge. Started by Doug Lenat in 1984, this project utilizes knowledge engineers who handcraft assertions and place them in Cyc's logical frameworks, using a logical representation called CycL. To use Cyc for natural language tasks, one must translate text into CycL through a complex and difficult process, as natural language is ambiguous while CycL is logical and unambiguous.

2.3.2 WordNet

Princeton University's WordNet [4] is one of the most widely used natural language processing resources today. WordNet is a collection of words arranged into a hierarchy, with each word carefully divided into distinct "senses" with pointers to related words, such as antonyms, *is-a* superclasses, and words connected by other relations such as *part-of*. WordNet's popularity may be explained by the ease a researcher has in interfacing it with a new application or system. We have endeavored to accomplish this flexibility of integration with ConceptNet.

2.3.3 BSO

Currently being developed, the Brandeis Semantic Ontology (BSO) [13] is a large lexical resource based in James Pustejovsky's Generative Lexicon (GL) [12], a theory of semantics that focuses on the distributed nature of compositionality in natural language. Unlike ConceptNet, however, the BSO focuses on the type structure and argument structure as well as on relationships between words.

An important part of GL is its network of qualia relations that characterize the relationships between words in the lexicon, and this structure is significantly similar to the set of ConceptNet relations. There are four types of qualia relations: *formal*, the basic type distinguishing the meaning of a word; *constitutive*, the relation between an object and its parts; *telic*, the purpose or function of the object; and *agentive*, the factors involved in the object's origins [12].

We've noticed that these qualia relations line up well with ConceptNet 3 relations. *IsA* maps well to the formal qualia, *PartOf* to the constitutive, *UsedFor* to the telic. The closest relation in ConceptNet 2 to the agentive relation was the *CapableOfReceivingAction* relation, but this is too general, as it describes many things that can happen to an object besides how it comes into being. In order to further this GL compatibility, we've added the *CreatedBy* relation and implemented targeted elicitation frames to collect statements that correspond with the agentive qualia.

3 The Design of ConceptNet 3

In developing ConceptNet 3, we drew on our experience with working with ConceptNet as users and observed what improvements would make it easier to work with. The new architecture of ConceptNet is more suitable to being incrementally updated, being populated from different data sources, and searching in complex queries such as those that are necessary to discover common-sense analogies. We believe that these improvements make ConceptNet more accessible to a variety of developers of artificial intelligence applications.

3.1 Concepts

The basic nodes of ConceptNet are *concepts*, which are aspects of the world that people would talk about in natural language. Concepts correspond to selected constituents of the common-sense statements that users have entered; they can represent noun phrases, verb phrases, adjective phrases, or prepositional phrases (when describing locations). They tend to represent verbs only in complete verb phrases, so "go to the store" and "go home" are more typical concepts than the bare verb "go".

Although they are derived from constituents, concepts are not literal strings of text; a concept can represent many related phrases, through the normalization process described later.

3.2 Predicates

In a semantic network where concepts are the nodes, the edges are *predicates*, which express relationships between two concepts. Predicates are extracted from the natural language statements that contributors enter, and express types of relationships such as IsA, PartOf, LocationOf, and UsedFor. Our 21 basic relation types are not a closed class, and we plan to add more in the future.

In addition to these specific relation types, there are also some underspecified relation types such as *ConceptuallyRelatedTo*, which says that a relationship exists between two concepts, but we can't determine from the sentence what it is. Though they are vague, these connections can help to provide information about the context around a concept, and they provide a fallback for cases where the parser is unable to parse a sentence. They are also used in several current applications [10].

Predicates maintain a connection to natural language by keeping a reference to the original sentence that generated them, as well as the substrings of the sentence that produced each of their concepts. This way, if the computer generates a new predicate without human input, like when it forms a hypothesis based on other knowledge, it can follow the example of other predicates to express this new predicate in natural language.

Relation	**Example sentence pattern**
IsA	NP is a kind of NP.
MadeOf	NP is made of NP.
UsedFor	NP is used for VP.
CapableOf	NP can VP.
DesireOf	NP wants to VP.
CreatedBy	You make NP by VP.
InstanceOf	An example of NP is NP.
PartOf	NP is part of NP.
PropertyOf	NP is AP.
EffectOf	The effect of VP is $NP \| VP$.

Table 1: *Some of the specific relation types in ConceptNet 3, along with an example of a sentence pattern that produces each type*

3.3 Modular Structure

ConceptNet 3 is built on top of the Common Sense Application Model of Architecture (CSAMOA) [1], a four-layer software design pattern intended to ease the development of common sense applications. By dividing components of common sense reasoning along consistent lines, CSAMOA encourages the development of reusable and interchangeable software components.

The layers of CSAMOA, in order, are the Corpus layer, which preserves original, human representation of common sense knowledge; the Representation layer, which abstracts the knowledge into a machine-interpretable form; the Realm layer, which helps navigate or performs generic computations on the structure of the machine-interpretable representation; and the Application layer, which is devoted to processing all user interactions and performing all other operations pursuant to the particular application. ConceptNet 3 was developed as a Representation layer for use with OMCS as the Corpus layer.

The use of CSAMOA and its emphasis on modularity represents a major change in the design choices underlying ConceptNet. In particular, we want the various components of ConceptNet, such as the parsing or reasoning components, to be customizable for different applications. For instance, the parsing patterns can be changed to handle different forms of natural language input, and the NLP procedures themselves can be replaced in order to generate a ConceptNet in a language besides English.

The most notable improvements CSAMOA has brought to ConceptNet are in its processing-oriented architecture. ConceptNet's data, data models, and processing code are now clearly separated, which permitted many advances in adding multiple language capabilities and improving the extraction of knowledge from unparsed text. ConceptNet's role in larger applications is also more clearly defined, allowing for the simplification of the code base.

4 Creating ConceptNet

4.1 Pattern Matching

Predicates in ConceptNet are created by a pattern-matching process, as they have been in previous versions [10]. We compare each sentence we have collected with an ordered list of patterns, which are regular expressions that can also include additional constraints on phrase types based on the output of a natural language tagger and chunker. These patterns represent sentence structures that are commonly used to express the various relation types in ConceptNet. Table 1 shows some examples of patterns that express different relations. The phrases that fill the slots in a pattern are the phrases that will be turned into concepts.

Many of these patterns correspond to elicitation frames that were presented on the OMCS website for users to fill in; the fact that so many sentences were elicited with predictable sentence structures means that these sentences can be reliably turned into predicates.

Other patterns, such as "NP is a NP", represent sentence structures that contributors commonly used when entering knowledge as free text. For these patterns, the constraints on phrase types (such as NP) imposed by the chunker are particularly important to prevent false matches.

Before a sentence goes through the pattern-matching process, common typographical errors and spelling mistakes are corrected using a simple replacement dictionary. If the sentence is a complex sentence with multiple clauses, we use patterns to extract simpler sentences out of it to run through the process.

4.2 Normalization

When a sentence is matched against a pattern, the result is a "raw predicate" that relates two strings of text. The *normalization* process determines which two concepts these strings correspond to, turning the raw predicate into a true edge of ConceptNet.

The following steps are used to normalize a string:

1. Remove punctuation.
2. Remove stop words.
3. Run each remaining word through a stemmer. We currently use Porter's Snowball stemmer, in both its English[1] and Portuguese versions [11].
4. Alphabetize the remaining stems, so that the order of content words in the phrase doesn't matter.

A concept, then, encompasses all phrases that normalize to the same text. As normalization often results in unreadable phrases such as "endang plant speci" (from "an endangered species of plant"), the normalized text is only used to group phrases into concepts, never as an external representation. This grouping intentionally lumps together many phrases, even ones that are only related by accidents of orthography, because we have found this to be an appropriate level of granularity for reasoning about undisambiguated natural language text collected from people.

[1] For compatibility with previous work, we use the original version of the English Snowball stemmer (the one commonly called "the Porter stemmer"), not the revised version.

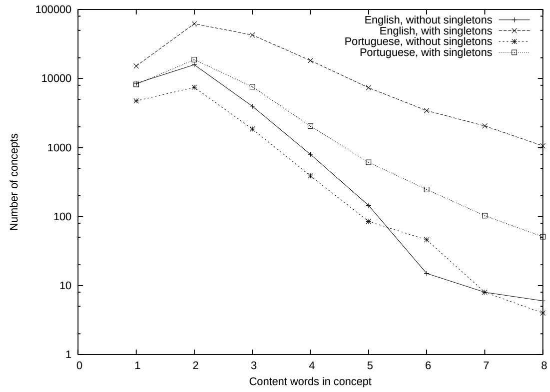

Fig. 1: *The number of words in the texts of concepts after normalization. The "without singletons" lines leave out sporadic concepts that only appear in one predicate, discarding many phrases that are too long to be useful concepts*

4.3 Open Mind Commons

ConceptNet would be nothing without the ability to collect knowledge from contributors on the Internet. The statements that currently comprise ConceptNet were collected from the Open Mind Common Sense web site, which used prompts such as "What is one reason that you would **ride a bicycle?**" to collect statements of common sense from its users.

Open Mind Commons [15] is an update of the original knowledge-collection website, OMCS 1, built on top of ConceptNet 3. The interface now includes activities that help refine its existing knowledge, by giving feedback to its users about what it already knows and what gaps seem to exist in its knowledge.

This feedback arises from a process that discovers analogies among the existing knowledge in ConceptNet. If concept X and concept Y appear in corresponding places in many equivalent predicates, they are considered to be similar concepts. Then, if concept X appears in a predicate that is not known about concept Y, Open Mind Commons can hypothesize that the same predicate is true for Y, and it can make this inference stronger by finding other similar concepts that lead to the same hypothesis. By following the links to natural language that are maintained in ConceptNet, it can turn the hypothesized predicate into a natural language question, which it asks to a user of the site.

Another kind of question that Open Mind Commons will ask based on analogy is a "fill in the blank" question: if it determines that it doesn't know enough predicates of a certain type about a concept, compared to what it knows about similar concepts, it will ask the

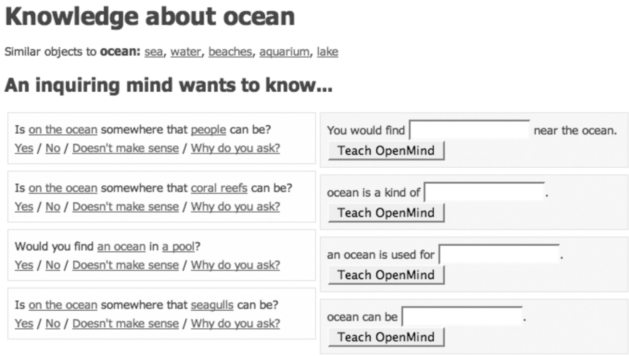

Fig. 2: *Open Mind Commons asks questions to fill gaps in its knowledge*

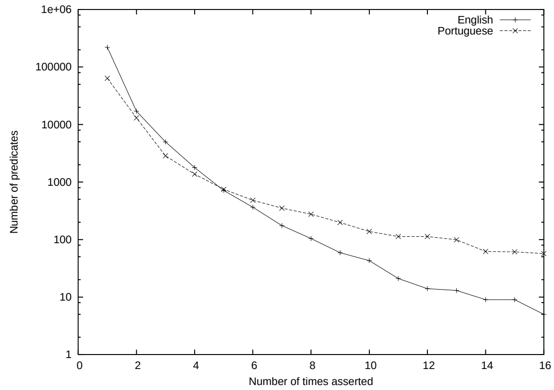

Fig. 3: *The distribution of scores among predicates extracted from OpenMind*

user to fill in that predicate. Figure 2 shows Commons asking both kinds of questions about the topic *ocean.*

Asking questions based on analogies serves to make the database's knowledge more strongly connected, as it eliminates gaps where simply no one had thought to say a certain fact; it also helps to confirm to contributors that the system is understanding and learning from the data it acquires.

4.4 Reliability of Assertions

In ConceptNet 3, each predicate has a score that represents its reliability. This score comes from two sources so far. A user on Open Mind Commons can evaluate an existing statement and increase or decrease its score by one point. The score can also be implicitly increased when multiple users independently enter sentences that map to the same predicate, and this is where the majority of scores come from so far.

The default score for a statement is 1—it is supported by one person: the person who entered it. Statements with zero or negative scores (because a user has decreased their score) are considered unreliable, and are not used for analogies in Open Mind Commons. Statements with positive scores contribute to analogies with a weight that scales logarithmically with their score.

In general, a significant number of predicates were asserted multiple times; Figure 3 shows the distribution of scores among all these predicates. Surprisingly, although the Portuguese corpus has been around for a shorter time and has fewer predicates, its predicates tend to have higher scores. The fact that all Portuguese statements were entered through structured templates, not through free text, may have caused them to coincide more often.

The highest-scored predicate in the English OMCS is "Dogs are a kind of animal", asserted independently by 101 different users. The highest-scored predicate in Portuguese is "Pessoas dormem quando elas estão com sono" ("People sleep when they are tired"), asserted independently by 318 users.

4.5 Polarity

In ConceptNet 3, we have introduced the ability to represent negative assertions. This capability allows us to develop interfaces that may ask a question of a user and draw reasonable conclusions when the answer is "no." The pattern matching process includes additional patterns, which match sentences expressing the negation of one of our relation types.

To this end, we added a *polarity* parameter to our predicate models that can assume the values 1 and -1, and we introduced a collection of extraction patterns that mirror most of our other extraction pattern but detect negation. About 1.8% of the English predicates and 4.4% of the Portuguese predicates currently in ConceptNet 3 have a negative polarity.

Importantly, score and polarity are independent quantities. A predicate with a negative polarity can have a high, positive score, indicating that multiple users have attested the negative statement (an example is "People don't want to be hurt"). Predicates with a zero or negative score, meanwhile, are usually unhelpful or nonsensical statements such as "Joe is a cat" or "A garage is for asdfghjkl", not statements that are "false" in any meaningful way.

5 Evaluation

The quality of the data collected by OMCS was measured in a 2002 study [14]. Human judges evaluated a random sample of the corpus and gave positive results, judging three quarters of the assertions to be "largely true", over four fifths to be "largely objective and sensible", and 84% "common enough to be known by someone by high school".

Here, we evaluate ConceptNet 3 in a different way: by testing how often its assertions align with assertions in similar lexical resources. The structure of Cyc is not readily aligned with ConceptNet, but WordNet and the BSO both contain information that is comparable to a subset of ConceptNet. In particular, certain

ConceptNet relations correspond to WordNet's pointers and the BSO's qualia, as follows:

ConceptNet	**WordNet**	**BSO**
IsA	Hypernym	Formal
PartOf	Meronym	Constitutive
UsedFor	*none*	Telic

BSO's fourth qualia type, Agentive, corresponds to the ConceptNet relation CreatedBy, but this relation is new in ConceptNet 3 and we have not yet collected examples of it from the public.

In this evaluation, we examine IsA, PartOf, and UsedFor predicates in ConceptNet, and check whether an equivalent relationship holds between equivalent entries in WordNet and the BSO. The test set consists of all predicates of these types where both concepts normalize to a single word (that is, they each contain one non-stopword), as these are the concepts that are most likely to have counterparts in other resources. Such predicates make up 11.1% of the UsedFor relations, 21.0% of IsA, and 31.2% of the PartOf relations in ConceptNet.

For each predicate, we determine whether there exists a connection between two entries in WordNet or the BSO that have the same normalized form (stem) and the appropriate part of speech (generally nouns, except that the second argument of UsedFor is a verb). This allows us to make comparisons between the different resources despite the different granularities of their entries. If such a connection exists, we classify the predicate as a "hit"; if no such connection exists between the corresponding entries, we classify it as a "miss"; and if no match is possible because a resource has no entries with one of the given stems, we classify it as "no comparison".

The criterion for determining whether "a connection exists" does not require the connection to be expressed by a single pointer or qualia. For example, the only direct hypernym of the first sense of "dog" in WordNet is "canine", but we want to be able to match more general statements such as "a dog is an animal". So instead, we check whether the target database contains the appropriate relation from the first concept to the second concept *or* to any ancestor of the second concept under the IsA relation (that is, the hypernym relation or the formal qualia). Under this criterion, ConceptNet's (IsA "dog" "anim") matches against WordNet, as "anim" is the Porter stem of "animal", WordNet contains a noun sense of "dog" that has a hypernym pointer to "canine", and a series of hypernym pointers can be followed from "canine" to reach a sense of "animal".

There are two major classes of "misses". Sometimes, a ConceptNet predicate does not hold in another resource because the ConceptNet predicate is unreliable, vague, or misparsed; on the other hand, sometimes the ConceptNet predicate is correct, and the difference is simply a difference in coverage. We have assessed a sample of 10 misses between ConceptNet and WordNet in Table 3, and between ConceptNet and the BSO in Table 4.

We ran this evaluation independently for IsA, UsedFor, and PartOf predicates, against each of WordNet and the BSO (except that it is not possible to evaluate UsedFor against WordNet). As a control to show that not too many hits arose from random noise, we also tested "randomized IsA predicates". These predicates were created by making random IsA predicates out of the shuffled arguments of the IsA predicates we tested, so that these predicates would express nonsense statements such as "soy is a kind of peninsula". Indeed, few of these predicates were hits compared to real ConceptNet predicates, even though IsA predicates are the most likely to match by chance. Table 2 presents the results, and Figure 4 charts the success rates for each trial (the ratios of hits to hits plus misses).

Resource	Type	Hit	Miss	No comparison
WordNet	IsA	2530	3065	1267
WordNet	PartOf	653	1344	319
WordNet	Random	245	5272	1268
BSO	IsA	1813	2545	2044
BSO	PartOf	26	49	2241
BSO	UsedFor	382	1584	3177
BSO	Random	188	4456	2142

Table 2: *The results of the evaluation. A "hit" is when the appropriate concepts exist in the target database and the correct relationship holds between them, a "miss" is when the concepts exist but the relationship does not hold, and "no comparison" is when one or both concepts do not exist in the target database*

Missed predicate	**Reason for difference**
Swordfish is a novel.	unreliable
Bill is a name.	WordNet coverage
Sam is a guy.	vague
(offensive statement)	unreliable
A gymnasium is a hall.	vague
Babies are fun.	misparsed
Newsprint is a commodity.	WordNet coverage
Biking is a sport.	WordNet coverage
Cats are predators.	WordNet coverage
Seeds are food.	WordNet coverage

Table 3: *A sample of ConceptNet predicates that do not hold in WordNet, with an assessment of whether the difference comes from unreliable/vague information in ConceptNet or a difference in coverage*

A Pearson's chi-square test of independence showed that the difference in the hit vs. miss distribution between the real predicates and the randomly-generated ones is statistically very significant, with $p < 0.001$ ($df = 1$) for each relation type. WordNet has $\chi^2 = 2465.3$ for IsA predicates and $\chi^2 = 1112.7$ for PartOf predicates compared to random predicates; the BSO has $\chi^2 = 1834.0$ for IsA, $\chi^2 = 159.8$ for PartOf, and $\chi^2 = 414.7$ for UsedFor compared to random predicates.

6 Discussion

As a resource, ConceptNet differs from most available corpora in the nature and structure of its content. Unlike free text corpora, each sentence of OMCS was entered by a goal-directed user hoping to contribute common sense, resulting in a wealth of statements that focus on simple, real-world concepts that often go unstated.

ConceptNet predicate	Reason for difference
A contest is a game.	BSO coverage
A spiral is a curve.	BSO coverage
A robot is a worker.	vague
A cookie is a biscuit.	regional
An umbrella is waterproof.	misparsed
A peanut is a legume.	BSO coverage
A hunter is a camper.	unreliable
A clone is a copy.	BSO coverage
The president is a liar.	unreliable
People are hairdressers.	unreliable

Table 4: *A sample of ConceptNet predicates that do not hold in the BSO, with an assessment of whether the difference is due to unreliable information in ConceptNet or a difference in coverage*

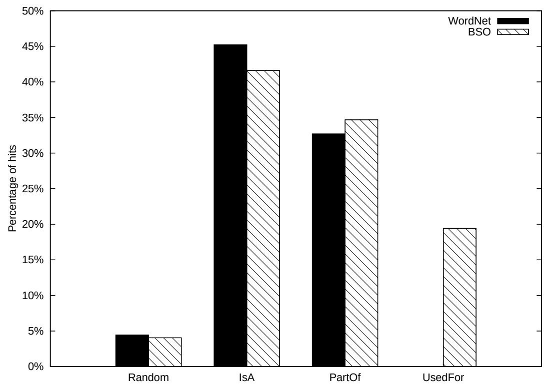

Fig. 4: *When ConceptNet predicates can be mapped onto relations between WordNet and BSO entries, they match a significant percentage of the time*

Our evaluation has shown that our information frequently overlaps with two expert-created resources, WordNet and the Brandeis Semantic Ontology, on the types of predicates where they are comparable. The goal of ConceptNet is not just to emulate these other resources, though; it also contains useful information beyond what is found in WordNet or the BSO. For example, many "misses" in our evaluation are useful statements in ConceptNet that simply do not appear in the other resources we evaluated it against, such as "sauce is a part of pizza", "a son is part of a family", and "weekends are used for recovery".

In addition, ConceptNet expresses many important types of relations that we did not evaluate here, such as CapableOf ("fire can burn you", "birds can fly"), LocationOf ("you would find a stapler on a desk", "you would find books at a library"), and EffectOf ("an effect of opening a gift is surprise", "an effect of exercise is sweating"). All of these kinds of information are important in giving AI applications the ability to reason about the real world.

References

[1] J. Alonso. CSAMOA: A common sense application model of architecture. *Proceedings of the Workshop on Common Sense and Intelligent User Interfaces*, 2007.

[2] J. Anacleto, H. Lieberman, M. Tsutsumi, V. Neris, A. Carvalho, J. Espinosa, and S. Zem-Mascarenhas. Can common sense uncover cultural differences in computer applications? In *Proceedings of IFIP World Computer Conference*, Santiago, Chile, 2006.

[3] H. Chung. *GlobalMind — bridging the gap between different cultures and languages with common-sense computing.* PhD thesis, MIT Media Lab, 2006.

[4] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge, MA, 1998.

[5] P. Grice. Logic and conversation. In *Speech Acts.* Academic Press, 1975.

[6] C. Havasi. An evaluation of the brandeis semantic ontology. *To Appear In Proceedings of the Fourth International Workshop on Generative Approaches to the Lexicon*, 2007.

[7] D. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 11:33–38, 1995.

[8] H. Lieberman, A. Faaborg, W. Daher, and J. Espinosa. How to wreck a nice beach you sing calm incense. *Proceedings of the 10th international conference on Intelligent user interfaces*, 2005.

[9] H. Lieberman, H. Liu, P. Singh, and B. Barry. Beating some common sense into interactive applications. *AI Magazine*, 25(4):63–76, 2004.

[10] H. Liu and P. Singh. ConceptNet: a practical commonsense reasoning toolkit. *BT Technology Journal*, 2004.

[11] M. F. Porter. Snowball: a language for stemming algorithms. Snowball web site, 2001. http://snowball.tartarus.org/texts/introduction.html – accessed Jan. 31, 2007.

[12] J. Pustejovsky. *The Generative Lexicon.* MIT Press, Cambridge, MA, 1998.

[13] J. Pustejovsky, C. Havasi, R. Saurí, P. Hanks, and A. Rumshisky. Towards a generative lexical resource: The Brandeis Semantic Ontology. *Proceedings of the Fifth Language Resource and Evaluation Conference*, 2006.

[14] P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu. Open Mind Common Sense: Knowledge acquisition from the general public. In *Proceedings of First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, Irvine, California, 2002.

[15] R. Speer. Open Mind Commons: An inquisitive approach to learning common sense. *Proceedings of the Workshop on Common Sense and Interactive Applications*, 2007.