Designing Interfaces for Guided Collection of Knowledge about Everyday Objects from Volunteers

Timothy Chklovski
Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292 USA
+1 310 448 8763
timc@isi.edu

ABSTRACT

A new generation of intelligent applications can be enabled by broad-coverage knowledge repositories about everyday objects. We distill lessons in design of intelligent user interfaces which collect such broad-coverage knowledge from untrained volunteers. We motivate the knowledge-driven template-based approach adopted in LEARNER2, a second generation proactive acquisition interface for eliciting such knowledge. We present volume, accuracy, and recall of knowledge collected by fielding the system for 5 months. LEARNER2 has so far acquired 99,018 general statements, emphasizing knowledge about parts of and typical uses of objects.

Categories and Subject Descriptors: I.2.6 [**Artificial Intelligence**]: Learning – *knowledge acquisition*; I.2.4 Knowledge Representation Formalisms and Methods: *frames and scripts, semantic networks*

General Terms: Algorithms, Design, Experimentation, Human Factors.

Keywords: Interfaces for knowledge elicitation, collecting broad-coverage knowledge repositories, generalization in knowledge acquisition

1. INTRODUCTION

Broad-coverage knowledge repositories containing extensive and detailed knowledge about everyday objects hold the promise of enabling a new generation of intelligent user interfaces and natural language understanding systems. Equipping computers with knowledge about objects in everyday world can potentially pave way to software assistants which can be more helpful with real-world tasks because they are knowledgeable about everyday objects and their uses. Natural language understanding tasks can also benefit from a broad-coverage resource of world knowledge. Usefulness of more extensive broad-coverage resources has been argued more extensively in [1],[9]. Currently, WordNet [10] is probably the most widely used broad-coverage semantic resource which has been used as source of both lexical and world knowledge. This resource's impact on the research community can be gauged by the size of the WordNet bibliography [13], numbering hundreds of uses in research. CYC [9] is another example of a broad-coverage resource. Both were created manually over many years by a small number of experts.

An alternate approach we are investigating is to create repositories of this kind by eliciting the knowledge from a large number volunteers [1],[2],[12]. This may be a viable, low-cost approach to creating dense, broad, up-to-date knowledge repositories.

LEARNER2, a web-based system for proactive acquisition of knowledge about everyday objects, has been deployed as a kiosk at an ongoing science museum exhibit visiting seven cities over three years and on the web at *http://learner.isi.edu*. LEARNER2 is a second generation system, implemented from scratch. LEARNER2 adopts an approach formulated as a result of collecting knowledge over the web for two years with an earlier system, LEARNER. During 5 months of collection, LEARNER2 acquired 99,018 entries. The collected knowledge includes knowledge about parts of objects, such as <*door, car*>, <*flag, mailbox*>, <*knob, door*> and knowledge about typical uses of objects, such as <*glasses, see*>, and <*fan, cool*>. To collect such knowledge, guidance and proactive prompting are used to focus the contributor on providing and elaborating the specific knowledge needed.

2. APPROACH TO COLLECTION IN LEARNER2

LEARNER2's approach to knowledge collection is based on lessons learned from prior work. We briefly overview this prior work and introduce the lessons extracted. LEARNER, previously developed by the author [1,2], relies on analogical reasoning over canonicalization of parsable natural language to formulate and proactively pose knowledge acquisition questions. More than 200,000 parsable stand-alone statements have been collected over two years from more than 4,000 contributors. LEARNER also solicited feedback on the experience of using the interface, and more than 200 comments have been received and reviewed to date. *Open Mind Common Sense* (OMCS) [12] is another project for collecting knowledge from volunteers. Over four years of deployment OMCS has collected more than 600,000 entries including stand-alone statements, image descriptions, story fragments, and paraphrases. OMCS allows contributors to enter knowledge as free-standing assertions and by filling in blanks in templates, but does not proactively formulate follow-on questions.

The end-to-end experience with LEARNER and close work with OMCS data have provided several lessons along the way. The lessons revolve around three aspects of the collected knowledge: *interpretability, density vs. breadth*, and *generalization*.

Interpretability. The difficulty of interpreting natural language precisely is well known. Experience with OMCS and LEARNER has highlighted just how insidious this difficulty is. OMCS contained loosely phrased templates which admitted ambiguous se-

IUI'05, January 10–13, 2005, San Diego, California, USA.
ACM 1-58113-894-6/05/0001.

mantic relations. The looseness of templates limited the interpretability of the collected knowledge. For example, the template "*a <action> is for* ______" has collected, without distinction, assertions in which *is for* stood for at least three different semantic relations: *results in a (emotional) state*, as in "riding a horse *is for* pleasure," *is carried out by (typical actor)*, as in "eating breakfast in bed *is for* sick people," and *allows application of*, as in "getting a job *is for* using your skills." Additionally, the template's phrasing admitted syntactically complex entries as arguments, further complicating processing.

LEARNER restricted the input it collected to only parsable sentences, and tended to limit occurrence of elaborate syntactic structures by seeding its collection be analogy with statements no longer than seven words. Still, the entries were structurally ambiguous. LEARNER also left the semantic relations within the sentences underconstrained, as they commonly are in natural language. The lesson we extracted from these observations was: the semantic relation and variation in the input need to be more tightly controlled to increase interpretability.

Density vs. Breadth. A collection effort may benefit from sequentially focusing on acquiring specific types of knowledge. Uses of WordNet in a variety of applications from processing user input to interpreting free text suggest that dense coverage of instances is important for a given relation to be useful. Because neither OMCS nor LEARNER focus acquisition on few relations (OMCS has more than 29 different collection "activities", each resulting in its own type of knowledge), the resources they create tend towards breadth (many types of relations with shallow coverage), rather than density. We noted that emphasizing density by focusing on one or two most important semantic relations would rapidly deliver a dense collection for these relations.

Generalization. A collection effort can benefit from generalizing across what it has learned to rapidly learn the scope of entities for which a given property holds (e.g. the property "*has a part called a* wing" holds for all birds). Fielding LEARNER revealed that it would occasionally focus on acquiring knowledge common to most physical objects, e.g. *X can be burnt in a fire* and *X can be viewed with a camera*, to the exclusion of learning about the more characteristic properties. The system did not generalize to learn about a broader class which included known instances. Similarly, OMCS did not generalize from knowledge it acquired. We continue to investigate conditions under which it is appropriate to generalize from collected knowledge.

LEARNER2 addresses *interpretability* by relying on carefully designed templates for eliciting knowledge. These templates constrain the semantic relation between template slots. We further address interpretability by adopting an *incremental acquisition* approach. With the view that learning happens on the fringes of knowledge and in piecemeal fashion, LEARNER2 builds on its knowledge a very small piece at a time. This is done by designing individual templates to discourage entry of complex constructs in a single template slot.

The second theme of *density vs. breadth* is treated in favor of focusing on density first, and is addressed by again relying on templates carefully designed to each collect a single semantic relation.

The third theme of generalization as part of knowledge acquisition is addressed by a broader mechanism of *invocation rules*. Each template comes with an invocation rule which determines when it can be posed and how it should be partially pre-filled with the knowledge just collected from the contributor and, if the invocation rule so specifies, with previously collected knowledge. This mechanism also enables incremental acquisition.

LEARNER2 has been deployed to collect knowledge about common real-world objects, focusing on information about their parts and their uses. In selecting the relations, we targeted two categories: relations *partially available* and *unavailable* in publicly accessible broad-coverage resources. Partially available relations were targeted to extend existing resources and provide a basis for comparison of coverage (recall) attained with our methods. We selected the meronymy (*part-of*) relation as partially available and potentially useful in many settings, ranging from diagnostic dialogues to cases of anaphora resolution (e.g., resolving *its* in "its wheels are missing"). For unavailable relations, we looked to commonly used relations linking two different parts of speech. We collect the relation specifying typical actions of objects, focusing on typical uses tools (e.g. "*a* copier *is typically used to* copy *something*") and typical patients of those uses ("*a* copier *is typically used to* copy documents"). These relations can be important to an intelligent system in deciding what tool to use to achieve a certain goal, inferring goals of agents seeking specific tools, and ensuring the patient objects (e.g. "documents") are available when supporting carrying out a wide range of tasks. The remaining types of collected relations build on the above by generalizing from parts to function (*part implies possible use*), and about similarity judgments about objects, useful in reasoning by analogy. When providing knowledge, contributors can mark a given question as "unreasonable," (e.g., "*a* bar of soap *has a piece or a part called* ________."), which allows to pose questions without needing a guarantee that an answer exists.

3. KNOWLEDGE COLLECTED

The kiosk deployment of LEARNER2 is a part of the "Robots and Us" traveling exhibit. The exhibit and kiosk have been up for 5 months at the Science Museum of Minnesota, having gathered the data presented here; the exhibit is set to visit six more locations in the coming 2.5 years, and data is also being gathered over the Web. So far, 99,018 entries have been collected. These were post-processed to discard or repair a variety entries which despite the guidance of the interface were still of "invalid" – including misspellings and results of hitting the keys randomly. This resulted in 75,071 (75.8% of all collected) entries[1] in which each argument contains only terms present in WordNet with the correct part of speech. The WordNet-based filtering was used to discard malformed entries and to ease using the resource together with WordNet. The number of entries of each type is shown in Table 1.

Table 1: **Number of entries in the output dataset, by relation**

Relation	Count	% of total
Meronymy	24,747	33.0%
Has typical use	24,714	32.9%
Patient of typical use	5,676	7.6%
Part implies possible use	5,858	7.8%
Similar	14,076	18.8%
Total	**75,071**	**100.0%**

To give an example of the knowledge collected, Table 2 shows all meronymy and typical-use-of relations collected about *telephone*.

[1] All data is available at *http://learner.isi.edu/downloads*

Table 2. Meronymy and typical use relations for "telephone"

Parts of Telephone		Typical use of Telephone
amplifier	microphone	call (7x)
base	mouth	communicate (6x)
cord	mouthpiece	hear (2x)
dial	speaker (2x)	listen
earpiece (3x)	telephone cord	phone
handle	wire (2x)	

The resulting resource contains entries which may be questionable or even incorrect. In a preliminary evaluation of 94 meronymy relations, 78% were judged "good" (e.g. <*engine, car*>, <*screen, television*>, <*zipper, backpack*>), 15% "questionable" (e.g. <*hole, basketball*>, <*head, beer*>, <*joint, bone*>, <*shaft, elevator*>), and the remaining 7% "clearly wrong" (e.g. <*mainframe, system*>, <*arm, machine*>). For 87 "has typical use" relations, 71% were judged "good" (e.g. <*bag, hold*>, <*bat, hit*>, <*cereal, feed*>, <*shed, store*>), 20% "questionable" (<*basketball, dribble*>, <*book, read*>, <*disc, spin*>), and 9% "clearly wrong" (e.g. <*kitten, cuddle*>, <*wall, push*>). In some cases, identifying word senses would be important. Here are some of the more extreme examples (judged "good"): *part-of*(wing, building), *part-of* (head, beer), *has-typical-use*(organ, play).

We restrict the (relative) recall analysis to the meronymy relation and to 326 objects used to seed the expanding set of objects used to instantiate templates into questions when starting afresh. We omit from the analysis entries with slots which are not WordNet entities, and disregard sense information. To date LEARNER2 collected 4,173 distinct assertions matching these restrictions. WordNet contains 1,352 assertions directly about parts of the 326 objects, and a total of 7,439 assertions if meronymy is looked up for hypernyms (e.g., a *truck* has an *airbrake*, a part of a *motor vehicle*). LEARNER2 recalled 304 (4.1%) of the 7,439 parts. Conversely, WordNet recalled 7.3% of this subset of LEARNER2 data.

4. DISCUSSION AND CONCLUSION

Using templates, rather than parsable natural language as in LEARNER not only addresses lessons learned (as discussed), but introduces the following benefits: frees users from generating fully grammatical sentences (which was at times challenging, according to user feedback), frees the system from generating fully grammatical sentences, and (together with invocation rules) allows generation of questions without requiring seed statements (needing only seed objects). However, the template-based approach also limits types of statements collected.

An alternative to collecting knowledge from volunteers is extracting it from the Web, including the WebFountain text analytics platform [6], KnowItAll [5], VERBOCEAN [3] and Hearst's original [7] or more uniform machine-readable corpora such as dictionaries, e.g. MindNet [4]. These approaches seek to identify lexico-syntactic patterns frequent in the source corpus. In our experience, it has been important to use carefully designed templates which tightly constrain the semantics of the entered statements. It is unlikely that many instances of such "overly precise" templates are present in corpora intended for human consumption, especially considering that the larger corpora tend to be less rigorously constructed. Still, further research is needed to assess the tradeoffs in terms of precision (correctness), coverage, and complexity of knowledge collected by this approach with knowledge extracted by state of the art text mining. Future work on collecting the knowledge can also be instructed by additional work on using the collected knowledge to construct intelligent applications.

Other related work looks at credit assignment and effort allocation for volunteer contributions [11] and theoretical results on amount of validation required under different noise levels [8].

We presented LEARNER2, a new system for collecting knowledge about everyday objects. This system addresses three themes in designing interfaces for collecting knowledge from volunteers importance of which was learned from two years of experience: (I) interpretability, (II) density vs. breadth, and (III) integration of generalization in the acquisition cycle. LEARNER2 addressed these with two features: carefully designed templates which admit syntactically simple individual arguments and control ambiguity of semantic relations, and mechanism of invocation rules which allow incremental collection and generalization of knowledge. This type of approach may be a viable and competitive way to create broad-coverage knowledge repositories about everyday objects, enabling a new generation of intelligent applications.

5. ACKNOWLEDGMENTS

We acknowledge funding for this work by DARPA under contract no. NBCHD030010. We also thank Joshua Seaver of the Science Museum of Minnesota for his work on deploying the kiosk.

6. REFERENCES

[1] Chklovski, T. *Using Analogy to Acquire Commonsense Knowledge from Human Contributors*, PhD thesis. MIT Artificial Intelligence Laboratory technical report AITR-2003-002, 2003

[2] Chklovski, T. LEARNER: A System for Acquiring Commonsense Knowledge by Analogy. In *Proceedings of Second International Conference on Knowledge Capture (KCAP)*, 2003

[3] Chklovski, T. and Pantel, P. VERBOCEAN: Mining the Web for Fine-Grained Semantic Verb Relations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-04)*. Barcelona, Spain, 2004

[4] Dolan, W., Vanderwende, L., and Richardson, S. D. Automatically deriving structured knowledge bases from on-line dictionaries. Technical Report MSR-TR-93-07, Microsoft Research, 1993

[5] Etzioni, O., Cafarella, M., Downey, D., Popescu, A., Shaked, T., Soderland, S., Weld, D., Yates, A. Methods for Domain-Independent Information Extraction from the Web: An Experimental Comparison, in *AAAI*, 2004

[6] Gruhl, D., Chavet, L., Gibson, D., Meyer, J., Pattanayak, P., Tomkins, A., and Zien, J. How to build a WebFountain: An architecture for very large-scale text analytics. *IBM Systems Journal.* 43(1):64–77, 2004

[7] Hearst, M. Automatic acquisition of hyponyms from large text corpora. *COLING-92*, pp. 539–545, 1992

[8] Lam, C. and Stork, D. Evaluating classifiers by means of test data with noisy labels, *IJCAI-2003*. pp. 513–518

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

[10] Miller, G. WordNet: An On-line Lexical Database. In *International Journal of Lexicography*, Vol.3, No.4, 1990

[11] Richardson, M., Domingos, P. Building large knowledge bases by mass collaboration, in *Proceedings of Second International Conference on Knowledge Capture (K-CAP 2003)*.

[12] Singh, P., Lin, T., Mueller, E., Lim, G., Perkins, T., Zhu, W. Open Mind Common Sense: Knowledge acquisition from the general public. In Robert Meersman & Zahir Tari (Eds.), LNCS: Vol. 2519. *On the Move to Meaningful Internet Systems: DOA/CoopIS/ODBASE* (pp. 1223-1237). Springer-Verlag 2002

[13] WordNet bibliography, engr.smu.edu/~rada/wn