

Building a General Knowledge Base of Physical Objects for Robots

Valerio Basile¹, Elena Cabrio², and Fabien Gandon¹

¹ INRIA Sophia Antipolis, France

{valerio.basile|fabien.gandon}@inria.fr

² University of Nice Sophia Antipolis, France

elena.cabrio@unice.fr

Abstract. In this paper we present an ongoing work on building a repository of knowledge about objects typically found in homes, their usual locations and usage. We extract an RDF knowledge base by automatically reading text on the Web and applying simple inference rules. The obtained common sense object relations are ready to be used in a domestic robotic setting, e.g. “a `frying pan` is usually located in the `kitchen`”.

1 Introduction

When working with and for humans, robots and autonomous systems must know about the objects involved in human activities. While great progress has been made in object instance and class recognition, a robot is always limited to knowing about the objects it has been trained to recognize. To overcome this issue, robots should be enabled to exploit the vast amount of knowledge on the Web to learn about previously unseen objects and to use this knowledge when acting in the real world. Methods to extract basic ontological knowledge about objects by analyzing unstructured and structured information sources on the Web are needed, to retrieve information about objects properties and functionalities.

In this paper, we address precisely this issue, that can be further divided into the following two subquestions: *i*) How much general knowledge is already present on the Web?, and *ii*) How to represent this knowledge in a way that robots can use for their real-world tasks?

In this paper, we propose a method based on “machine reading” [2] to extract formally encoded knowledge from unstructured text. Given the robot domestic setting scenario, we focus on extracting the following information: *i*) the type of the object, *ii*) where it is typically located, and *iii*) common semantic frames involving the object. Existing resources as ConceptNet or Cyc¹ attempt to assemble an ontology and knowledge base of everyday common sense knowledge, but these models are not sufficient to meet our objectives since they do not contain extensive object knowledge or the right level of abstraction. ConceptNet, for instance, provides for the `AtLocation`, `knife` relation of the concept `knife` a list of 34 candidates, including rooms and containers (`in_kitchen`, `backpack`) but

¹ <http://conceptnet5.media.mit.edu/>, <http://www.opencyc.org/>

also other kinds of locations, not necessarily prototypical (`pocket`, `oklahoma`, `your_back`). OpenCyc, on the other hand, contains a great deal of taxonomical knowledge about the entity `knife`, but nothing about its possible locations.

Our approach combines linguistic and semantic analysis of natural language with entity linking and formal reasoning to create a meaning bank of common sense knowledge. We leverage curated resources (e.g. DBpedia², BabelNet [8]) and learn from the unstructured Web when a knowledge gap occurs.

2 Information Extraction from Natural Language

To extract useful knowledge from unstructured data on the Web, two main types of language analysis are needed: semantic parsing (the process of extracting some formal representation of the meaning of the target text) and entity linking. Concerning the first one, we are mostly interested in *thematic roles*, i.e. the role played by a specific entity in a given event. We use the C&C tools pipeline plus Boxer [5, 3], in a similar way to the FRED information extraction API [9]. For the latter analysis, we use Babelify [7], an online API that, given an arbitrary text, returns a list of the word senses, represented as BabelNet synsets, and the URIs of DBpedia entities linked to the words that mention them.

Once the system has extracted the entities, properly linked to the LOD cloud, and decided the roles they play in the situations they are involved in, the only missing piece is a formal description of the situations themselves. For this, we resort to frame semantics [6], a theory of meaning that describes events and situations, along with the possible roles for each involved entity. FrameNet [1] provides a mappings to other linguistic and semantic resources, therefore we are able to link the events in the meaning representation returned by the semantic parsing stage to FrameNet frames. By integrating all these sources of information, from a text like “Annie eats an apple”, we extract RDF triples such as: `dbp:Apple vn:Patient fn:Ingestion`.³ Since these resources are already in the LOD (FrameNet relations are modeled through LEMON⁴), this method produces new triples that can be directly published back to the Web.

To automatically augment the knowledge base, we apply inference rules to the triples to extract new information on top of the language analysis. We exploit the relation of *co-mention* (i.e., two entities being mentioned together in a sentence or similar context) together with the DBpedia type hierarchy, to define the rule: if X is subject of a DBpedia category that is subsumed by `dbp:Tool` and Y is a category that is subsumed by `dbp:Room` and the two entities are co-mentioned, then a likely location for X is Y ($isTool(X) \wedge isRoom(Y) \wedge comention(X, Y) \Rightarrow location(X, Y)$). For example: `dbp:Knife` is `purl:subject`⁵ of `dbp:Category:Blade_weapons`, which is `skos:narrower`⁶ than `dbp:Category:Tools`. At the

² <http://dbpedia.org>

³ Prefixes `dbp`, `vn`, `fn` respectively link back to DBpedia, VerbNet and FrameNet.

⁴ <http://lemon-model.net/lexica/uby/fn/>

⁵ <http://dublincore.org/documents/dcml-terms/>

⁶ <http://www.w3.org/TR/swbp-skos-core-spec>

same time, `dbp:Kitchen` is `purl:subject` of `dbp:Category:Rooms`. Combining this information with the fact that `dbp:Knife` is co-mentioned with `dbp:Kitchen`, we infer that the likely location of a knife is the kitchen.

Evaluation. To test the quality of the created knowledge base, we applied the system to text from the Web and evaluated its output. We created two corpora of written English: *i*) the first comprises five short documents (952 words in total) extracted from the documentation of the RoCKIn@Home challenge⁷, that describe typical tasks for a domestic robot; *ii*) then, to learn common knowledge about objects in general, we also experimented with open domain text. Language learners material is a fit option, since it is usually made of short, simple sentences about concrete, day-to-day situations. The ESL YES website⁸ contains a collection of 1,600 free short stories and dialogues for English learners. We extracted and converted to plain text 725 short stories (83,532 tokens).

Running the pipeline on the two corpora, we obtained two knowledge bases composed of co-mentions and semantic role triples: 3,184 triples from RoCKIn@home (57 of which semantic roles) and 49,165 from ESL YES (2,953 semantic roles). 91.2% of semantic role triples from RoCKIn@home and 76.4% from ESL YES are not aligned with any frame in FrameNet. Next we apply the rule defined in Section 2 to extract location relations between tools and rooms. We implemented the rule in CoReSe [4] obtaining 5 location relations from RoCKIn@home (5 objects in 2 rooms) and 101 from ESL YES (49 objects in 14 rooms).

We manually inspected the set of location relation extracted: of the five `<objects, room>` pairs extracted from RoCKIn@home, two are definitely accurate, two are disputable, depending on the situation, and one is too generic, thus not very informative. Following the same methodology, out of 101 location relations extracted from the ESL YES corpus, we judged 42 of them correct (e.g., `dbp:Frying_pan, dbp:Kitchen`), 31 questionable (e.g., `dbp:Suitcase, dbp:Bathroom`), and 28 not informative, i.e. they contain the entities `dbp:Tool` or `dbp:Room` (these can be easily filtered out).

Discussion. A great amount of text is needed to build a large scale knowledge base. Several steps in the processing pipeline introduce mismatches, especially with respect to the coverage of the lexical resources involved, ultimately resulting in the extraction of relatively few triples, considering the size of the input text. On average, we extract roughly a little less than one semantic role triple with a proper frame per sentence. This issue can be circumvented by adding more text from the Web. However, the problem remains to find and retrieve text that contains the right kind of information, i.e., common knowledge about objects.

Another issue highlighted by this work is the difficulty of categorizing DBpedia entities, especially entries that are not named entities, such as the household items of our knowledge base. Inference rules like the one we defined are dependent on the correct classification of entities (in tools and rooms, in our case), therefore a high-coverage taxonomy of objects is needed.

⁷ http://rockinrobotchallenge.eu/RoCKIn_D1.1_short.pdf

⁸ <http://www.eslyes.com/>

Finally, a unified tokenization does not solve all the alignment problems. Our system maps discourse referents to entity URIs or FrameNet frames, but these in turns rely on the respective mappings between text and semantics made by Boxer and Babelfy, which are not always aligned.

3 Conclusions and Future Work

This paper presents a novel method to extract information from natural language text based on the combination of semantic parsing and entity linking. The knowledge extracted is represented as RDF triples, to exploit other Web resources to augment the result of the information extraction (e.g. we infer location relations between objects and rooms).

The availability of large quantities of text is crucial to build a large common sense knowledge base following our approach. However, not all source text is the same, as the outcome is influenced by genre, style, and most important, the topics covered in the text. As future work we aim at building a corpus similar to the language learners one, but orders of magnitude larger. Moreover, besides the issues we found with ConceptNet, its sheer size and variety of semantic links makes for a valuable resource that we plan to incorporate in our work.

References

1. C. F. Baker, C. J. Fillmore, and J. B. Lowe. The berkeley framenet project. In *Procs of the 36th Annual Meeting of ACL*, ACL '98, pages 86–90, 1998.
2. K. Barker, B. Agashe, S.-Y. Chaw, J. Fan, N. Friedland, M. Glass, J. Hobbs, E. Hovy, D. Israel, D. S. Kim, R. Mulkar-Mehta, S. Patwardhan, B. Porter, D. Tecuci, and P. Yeh. Learning by reading: A prototype system, performance baseline and lessons learned. In *Procs of the 22Nd National Conference on AI - Volume 1*, AAAI'07, pages 280–286. AAAI Press, 2007.
3. J. Bos. Wide-Coverage Semantic Analysis with Boxer. In J. Bos and R. Delmonte, editors, *Semantics in Text Processing. STEP 2008 Conference Procs*, volume 1 of *Research in Computational Semantics*, pages 277–286. College Publications, 2008.
4. O. Corby, R. Dieng-kuntz, and C. Faron-zucker. Querying the semantic web with the corese search engine. 2004.
5. J. Curran, S. Clark, and J. Bos. Linguistically motivated large-scale nlp with c&c and boxer. In *Proceedings of the 45th Annual Meeting of ACL Companion (Demo and Poster Sessions)*, pages 33–36, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
6. C. Fillmore. Frame semantics. *Linguistics in the morning calm*.
7. A. Moro, A. Raganato, and R. Navigli. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TAACL)*, 2:231–244, 2014.
8. R. Navigli and S. P. Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193(0):217 – 250, 2012.
9. V. Presutti, F. Draicchio, and A. Gangemi. Knowledge extraction based on discourse representation theory and linguistic frames. In *Procs of EKAW 2012*.