APPLYING NLP FOR BUILDING DOMAIN ONTOLOGY: FASHION COLLECTION

Inna Novalija, Gregor Leban
Artificial Intelligence Laboratory
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773144; fax: +386 1 4251038
e-mail: inna.koval@ijs.si

ABSTRACT

This paper presents an approach to developing a fashion domain ontology based on inputs from fashion experts and natural language processing (NLP) methods. While many of software solutions for fashion industry are concentrated on the design, manufacturing and trading applications, semantic technologies are just starting to interact with fashion domain. Domain ontologies allow capturing, sharing, analyzing and reusing the important information from the defined field.

1 INTRODUCTION

Ontologies are considered one of the pillars of Semantic Web and Semantic Technologies [1]. Gruber [2] defined Ontology as an explicit specification of a conceptualization consisting of the following main components: concepts, relations, functions, axioms and instances. Furthermore, ontologies enable effective domain knowledge representation, knowledge sharing and knowledge reuse [3]. Usage of ontologies allows to effectively discover patterns, by searching not only within the terms occurring in the query, but also within their semantically related concepts.

One example of large common-sense ontology is the Cyc Knowledge Base [4], which has been being developed for more than 20 years (more than 900 human years of effort) and is used as a knowledge source in the Cyc Artificial Intelligence system. It already aggregates more than 15.000 predicates, 300.000 concepts and 3.500.000 assertions.

Domain ontology are built upon knowledge from a particular domain.

While there are many semantic tools in some domains, such as biomedical, software engineering domains, other domains are just starting to interact with semantic technologies in general and ontologies in particular.

For instance, BioPortal [5] is web portal developed by the National Center for Biomedical Ontology (NCBO) [6], which provides access to number of resources (ontologies, terminologies, mappings), tools (ontology recommender, ontology annotator) and web services in biomedical domain. BioPortal users obtain a possibility of knowledge sharing and reuse in different knowledge representation formats, such as Web Ontology Language (OWL) and Open Biological and Biomedical Ontologies (OBO). BioPortal contains around 260 ontologies [5] from different groups. Unlike biomedical resources, in fashion domain there are no available semantic tools and formalized knowledge materials.

In this paper we present a methodology for development of the domain ontology in fashion domain.

As stated by Pearson [7], fashion is often at the forefront of technology usage. With technology development quickly accelerating, the fashion industry sees the convergence of nanotechnology, biotechnology, information technology and cognitive technologies.

The aim of this work is, with an assistance of sematic technologies, to create supportive mechanisms and tools, contributing to the improvement of information analysis and sharing processes both on the production and consumption sides of fashion industry.

The development of fashion ontology is meant to

- provide advanced search functionalities for fashion related content
- track what is going on in the fashion world
- show fashion entities related to each other.

The paper is structured as follows: Section 2 contains the related work on ontology learning; Section 3 describes the methodology for domain ontology development; Section 3 provides the insights into the created fashion ontology resources; finally, Section 4 concludes the paper.

2 RELATED WORK

Automatic or semi-automatic ontology learning based on text mining usually starts with corpus definition, collection and preprocessing [8]. Many ontology learning approaches are based on the expertise of domain experts, who validate concepts and relationships discovered from text. Grobelnik

and Mladenic [9] express the opinion that the process of ontology learning from text is closely connected to domain understanding and data understanding.
Natural language processing within ontology development problem has been covered by a number of researchers[10, 11].
Suchanek et al. [12] created a SOFIE system for ontology learning and population based on natural language document parsing and logical reasoning for disambiguation.
OntoGen [13] by Fortuna et al. is a system for topic ontology construction, which uses the vector-space model for document representation and operates based on a cosine similarity between textual documents.
A number of ontology extension and population methods based on lexico-syntactic patterns for ontology learning include Text2Onto [14] and SPRAT [15].
In comparison to the related work, the approach presented in this paper is heavily based on domain experts inputs required for fashion concepts extraction. The suggested methodology applies natural language processing and dependcy analysis for relation extraction.

3 APPROACH TO DOMAIN ONTOLOGY DEVELOPMENT

In this paper we propose an approach to creating domain specific ontology, based on user provided concept seeds for a particular domain. The approach consists of the phases described below:

1. **Collection and definition** of concept seeds.
2. **Mapping** seeds to **Wikipedia** and extending the ontology with relevant related concepts.
3. **Definition of relationships** between concepts.
4. Ontology **refinement**.

In the **first phase** the experts define a number of concepts and entitites for the particular domain.

Table 1: Fashion Seeds

Seeds by Classes	Number of Entities
Designer	650
Model	448
Clothing term	59
Trend	41
Season	76
Celebrity	383

In case of fashion domain, which was explored in this research, a team of fashion experts provided a list of fashion related concepts and entitites of several types: Designer, Model, Clothing term, Trend, Season, Celebrity.

Table 1 contains a fashion seeds statistics for each entity type.
Table 2 provides examples of seeds provided by fashion expers for each entity type.

Table 2: Examples of Fashion Seeds

Class	Example
Designer	Alexander McQueen
Model	Ava Smith
Clothing term	Hoodie
Trend	Safari
Season	Fall 2011 Womenswear
Celebrity	Penélope Cruz

The **next phase** of our approach constitutes mapping concept seeds to Wikipedia articles, obtaining more information about ontology concepts and new potentially related concepts. Mapping to Wikipedia is an automatic process, where concept seeds are compared to Wikipedia article titles.
Linking to Wikipedia provides us not only with textual description of a particular concept, but also allows obtaining structured information, such as Yago, Freebase and DBpedia inputs.

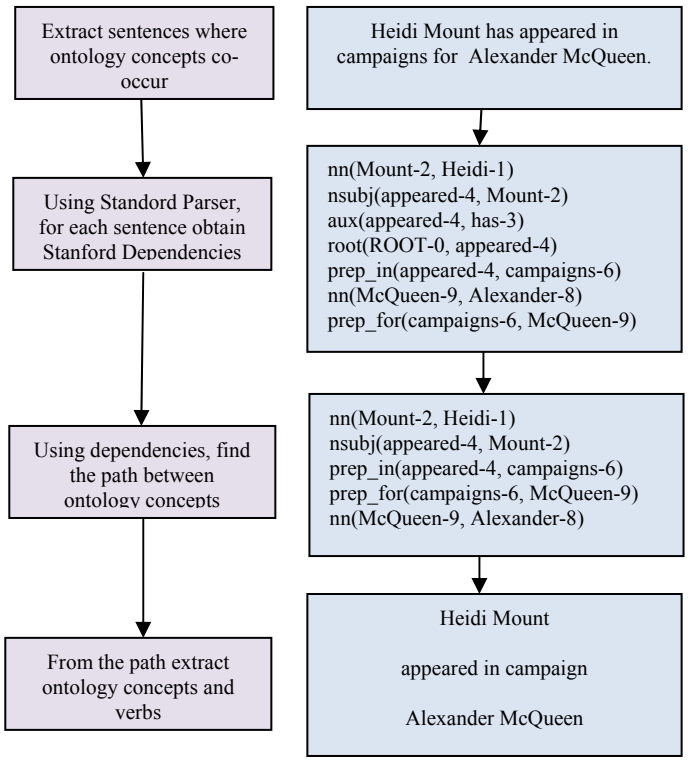

Figure 1: Example of relation extraction

The definition of a term usually can be found in the initial paragraph of Wikipedia article about this term. At the same time, in the first paragraph Wikipedia often provides links to the related terms, which can be explored as potential ontology concepts.

Definition of the relations is an important step of ontology building, since relations allow expanding the usage of our ontology.
While we can determine a general relation "linksTo" already from the Wikipedia links, more specific domain relations can be discovered with natural language processing (NLP) techniques.
For instance, there is a statement about model Heidi Mount:

> "Heidi Mount has appeared in campaigns for Alexander McQueen."

This statement contains a number of ontology concepts: Heidi Mount (*Model*), Alexander McQueen (*Designer*). Using NLP we can also identify a potential ontology relation: *Model appearedInCampaign Designer*:

- Heidi Mount,appeared in,Alexander McQueen

We suggest the following method for ontology relations detection based on Stanford parser [16] and Stanford Part-Of-Speech (POS) tagger [17] (see Figure 1). The suggested approach to relation extraction is based on analysis of dependencies between words in the sentence. The path between concepts is obtained through dependencies and correspondent verbs are extracted from the path.
More examples of relationships in fashion domains include:

- Model walkedInCampaign Designer
- Model modeledInCampaign Designer
- Designer created ClothingTern
- Celebrity appearedInCloth Designer etc.

In order to group synonymic relationships WordNet [18] is used to obtain verb synsets.

Ontology **refinement** follows the ontology learning and population process. At this stage it is important to keep only fashion appropriate concepts in the fashion ontology. Ontology refinement can be performed using a pool of websites related to fashion and checking if concepts from our ontology are mentioned at fashion websites. All non-frequent and non-relevant concepts are removed from fashion ontology.

4 FASHION ONTOLOGY

The current version of the generated ontology contains around 15.000 concepts and is published in the Resource Description Framework (RDF) format. In Example 1 we show the representation for fashion ontology entity "Heather Marks".

Example 2 demonstrates the representation of fashion entity "Revlon".

Example 1: RDF Representation for Fashion Entity "Heather Marks"

```
<rdf:Description rdf:about="http://ailab.ijs.si/fashion/resource/35481">
  <rdfs:label>Heather Marks</rdfs:label>
  <rdf:type rdf:resource="http://ailab.ijs.si/fashion/upperclass/Model"/>
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Person"/>
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Model"/>
  <rdf:type rdf:resource="http://dbpedia.org/class/yago/LivingPeople"/>
  <rdf:type rdf:resource="http://dbpedia.org/class/yago/CanadianFemaleModels"/>
  <rdf:type
  rdf:resource="http://dbpedia.org/class/yago/PeopleFromGreenwichVillage,NewYork"
  />
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/5538"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/49678"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/35002"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/14130"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/11294"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/1906"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/18121"/>
  <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/35481"/>
</rdf:Description>
```

Example 2: RDF Representation for Fashion Entity "Revlon"

```
<rdf:Description rdf:about="http://ailab.ijs.si/fashion/resource/5538">
 <rdfs:label>Revlon</rdfs:label>
 <rdf:type rdf:resource="http://dbpedia.org/ontology/Company"/>
 <rdf:type rdf:resource="http://dbpedia.org/ontology/Organisation"/>
 <rdf:type
rdf:resource="http://dbpedia.org/class/yago/CompaniesEstablishedIn1932"/>
 <rdf:type
rdf:resource="http://dbpedia.org/class/yago/CompaniesBasedInNewYorkCity
"/>
 <rdfs:comment>Revlon is an American cosmetics, skin care, fragrance, and
personal care company founded in 1932.</rdfs:comment>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/37731"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/48778"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/49826"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/11461"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/10692"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/13835"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/18127"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/639"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/35481"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/41994"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/49147"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/24085"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/4119"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/21751"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/5538"/>
 <ailab:linksTo rdf:resource="http://ailab.ijs.si/fashion/resource/31080"/>
</rdf:Description>
```

As it is visible from the examples that "Heather Marks" is an entity of type Model. In addition, we have obtained a number of types from DBpedia and Yago: Person, CanadianFemaleModels, PeopleFromGreenwichVillageNewYork. "Revlon" is a Company, Organisation, CompanyBasedInNewYorkCity, CompanyEstablishedIn1932. In addition, it is possible to see that "Heather Marks" is connected to "Revlon".

5 CONCLUSION AND FUTURE WORK

In this paper we presented an approach to developing a fashion domain ontology based on domain experts input and natural language processing methods. Fashion domain ontology allows capturing, sharing, analyzing and reusing the important information from the field of fashion.
The future work will include the improvements of relation extraction and ontology refinement methods, as well as creating semantically grounded applications in fashion domain.

6 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency, the IST Programme of the EC under PASCAL2 (IST-NoE-216886).

References

[1] Semantic Web Ontology, http://semanticweb.org/wiki/Ontology (accessed August 2013).
[2] Gruber, T. R. A translation approach to portable ontologies. Knowledge Acquisition 5/2, 199-220 (1993).
[3] Chandrasekaran, B.; Josephson, J. R.; Benjamins, R. V. What are Ontologies and why do we need them? In: IEEE Intelligent Systems and Their Applications 14, 20-26 (1999).
[4] Cycorp., Inc. , http://www.cyc.com (accessed August, 2013).
[5] Whetzel P.L., Noy N.F., Shah N.H., Alexander P.R., Nyulas C., Tudorache T., Musen M.A. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res. 2011 Jul; 39 (Web Server issue):W541-5. Epub 2011 Jun 14
[6] Musen M.A., Noy N.F., Shah N.H., Whetzel P.L., Chute C.G., Story M.A., Smith B.; NCBO team. The National Center for Biomedical Ontology. J Am Med Inform Assoc. 2012 Mar-Apr; 19(2): 190-5. Epub 2011, Nov 10.
[7] Pearson, I. The Future of Fashion, The Journal of The Communications Network (2005/4), http://www.podkolinski.co.uk/wordpress/wp-content/uploads/2012/01/fashion2.pdf
[8] Reinberger, M.-L.; Spyns, P. Unsupervised Text Mining for the Learning of DOGMA-Inspired Ontologies. In: Buitelaar, P.; Handschuh, S.; Magnini, B. (eds.) Ontology Learning from Text: Methods, Evaluation and Applications (IOS Press, 2005).
[9] Grobelnik, M.; Mladenić, D. Knowledge Discovery for Ontology Construction. In: Davies, J.; Studer, R.; Warren, P. (eds.) Semantic Web Technologies: Trends and Research in Ontology-Based Systems (John Wiley & Sons, 2006).
[10] Burkhardt, F.; Gulla, J. A.; Liu, J.; Weiss, C.; Zhou, J. Semi Automatic Ontology Engineering in Business Applications. In: Proceedings of Workshop Applications of Semantic Technologies, INFORMATIK 2008 (2008).
[11] Sabrina, T.; Rosni, A.; Enyakong, T. Extending Ontology Tree Using NLP Technique. In: Proceedings of National Conference on Research & Development in Computer Science REDECS 2001 (2001).
[12] Suchanek, F. M.; Sozio, M.; Weikum, G. SOFIE: A Self-Organizing Framework for Information Extraction. In: Proceedings of the 18th World Wide Web Conference, www2009 (Madrid, Spain, 2009).
[13] Fortuna, B.; Grobelnik, M.; Mladenić, D. OntoGen: Semi-automatic Ontology Editor. HCI 9, 309-318 (2007).
[14] Cimiano, P.; Völker, J. Text2Onto A Framework for Ontology Learning and Data-driven Change Discovery. In: Proceedings of NLDB 2005, 227-238 (2005).
[15] Maynard, D.; Funk, A.; Peters, W. SPRAT: a tool for automatic semantic pattern-based ontology population. In: Proceedings of International Conference for Digital Libraries and the Semantic Web (Trento, Italy, 2009).
[16] Stanford Parser, http://nlp.stanford.edu/software/lex-parser.shtml (accessed August 2013).
[17] Standofd Part-Of-Speech Tagger http://nlp.stanford.edu/software/tagger.shtml (accessed August 2013).
[18] WordNet, http://wordnet.princeton.edu (accessed August 2013).