Keynote speech

Nathalie Aussenac-Gilles

http://www.irit.fr/~Nathalie.Aussenac-Gilles/index_en.php

Semantic relations: new challenges in a world of linked data

Today, connecting two entities with a labelled, meaningful relation is a task that is required in such a large variety of situations that there are thousands of papers, research works and approaches about relation identification. Depending on the application using or producing "relations" or "links", this notion covers actually different realities:
from a natural language phrase to a formal formula, from a relationship between 2 entities to one between classes, from domain specific to universal relations, from contextual and possible relations to validated and proved links ... Terminology is one of these research and application domains where relations play a major role to structure and organise language. They contribute to account for meaning and illustrate the difficulty to fix a dynamic and moving reality in a static model.

In my talk, I shall first draw an overview of these diverse realities behind the notion of relation in research areas close to linguistics, terminology, knowledge modelling, natural language processing and semantic web. Then I shall connect these definitions with tools, methods and approaches to identify these relations. I shall finally comment how terminology can play a key role in this context, either to change its practices by integrating results and tools from other research fields, or by contributing to improve these approaches with the studies carried out during the last 20 years to build terminologies and terminological knowledge bases.

Detection of the Disease-Symptom Relation in Medical Texts in Catalan: a Corpus-Based Approach

Experts in terminology state that in specialized texts there are many terms connected among them, which form nodes and are distributed in conceptual structures. Following this idea, the study of conceptual and semantic relations in specialized texts is relevant. Specifically, the automatic or semiautomatic detection of these relations in real texts is a great challenge. In medicine, this challenge is especially important, since currently the developing of this kind of systems is useful for automatic construction and updating of ontologies, thesaurus, and definitions included in medical dictionaries and specialized manuals.

To carry out this kind of research, two main strategies are used: strategies based on linguistic patterns (e.g. Feliu, 2004) and machine learning strategies (e.g. Hearts, 1999; Riloff & Jones, 1999; Turney, 2006). Machine learning strategies are useful when a big annotated corpus is available. In Catalan there is a lack of some Natural Language Processing (NLP) tools. Specifically, there is not any annotated corpus with conceptual relations among terms and any system for detecting conceptual relations. The only work about this subject is the research by Feliu (2004), where she describes several linguistic patterns that would allow detecting some specific conceptual relations in specialized texts in Catalan.

In this context, our work has two main goals:

- to elaborate a list of linguistic patterns that show the disease-symptom relation between terms extracted from medical texts in Catalan,
- to propose a methodology for the automatic creation of these linguistic patterns.

The methodology that we propose consists of several stages. First, to select corpus including specialized texts in Catalan from the medical domain. Second, to choose some specific diseases to work with. Third, to extract contexts automatically from the corpus. Fourth, to analyse manually contexts including terminological units that express symptoms, in order to find explicit linguistic marks showing the disease-symptom relation. Fifth, by using these linguistic marks, to produce a list of general and lemmatized linguistic patterns. This list could be used by an automatic or semiautomatic system to detect this specific relation among terms in medical texts. In our research, we use the medical subcorpus of the IULA Technical Corpus and the online interface BwanaNet (Vivaldi, 2009) to extract the contexts automatically. Also, in our work, neurodegenerative diseases are chosen, taking into account that the study of their symptoms has a social interest, since their early diagnosis is nowadays necessary.

This research has obtained three main results:

- A list of 24 linguistic patterns that are used in medical texts in Catalan to express the disease-symptom relation are obtained, by using the proposed methodology.
- 31 different terms expressing symptoms of neurodegenerative diseases have been found in our corpus. From them, only the 22% appears in the definitions of these diseases in specialised dictionaries and manuals. For example, the term *estrès oxidatiu* ["oxidative stress"] is a symptom of the disease *esclerosi lateral amiotròfica* ["amyotrophic lateral sclerosis"]. We have found this term in our corpus, but it is not possible to find it in the definitions of *esclerosi lateral amiotròfica* included in the consulted terminological resources. This means that Specialized Texts Processing (STP) is an important tool for updating medical resources.
- Although in the terminology field it is stated that terms are the prototypical units that show specialized knowledge in texts, we have found that, in our corpus, symptoms are not only expressed by means of monolexical and polilexical terms, but also by means of specialized phraseology or collocations. For example, the disease *Alzheimer* ["Alzheimer disease"] is associated with the symptoms *estrès oxidatiu* ["oxidative stress"] (a prototypical terminological unit), but also with *aparició de dipòsits de plaques amiloides* ["appearance of beta-amyloid deposits"] (a specialized collocation formed by the term *plaques amiloides* and the collocative *aparició de dipòsits de.* Our research shows that, in our corpus, the 42% of symptoms are expressed by terms and the 58% by collocations.

The results obtained in this work are promising, because the list of generated linguistic patterns will be the basis for the implementation of a system to detect the disease-symptom relation. Also, the proposed methodology will allow us to generate more lists of patterns for the detection of other different conceptual relations in the medical field. Moreover, results are interesting because they help to understand how medical information is shown in specialized texts, regarding terms expressing new symptoms in texts, and symptoms expressed by collocations and not only by prototypical terminological units.

References

Feliu, J. (2004). *Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. [PhD. Thesis]

Hearst, M. A. (1992). «Automatic acquisition of hyponyms from large text corpora». In *Proceedings of COLING*. 539-545.

IULA Technical Corpus [on-line]. Barcelona: Institut Universitari de Lingüística Aplicada. http://bwananet.iula.upf.edu/

Riloff, E.; Jones, R. (1999). «Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping». In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*. 474-479

Turney, P.D. (2006). «Similarity of semantic relations». In *Computational Linguistics*, 32(3). 379-416.

Vivaldi, J. (2009). «Corpus and exploitation tool: IULACT and bwanaNet». In Cantos, Pascual; Sánchez, Aquilino (ed.). *A survey on corpus-based research*. Murcia: Asociación Española de Lingüística de Corpus. 224-239.

Topic words and content words: a statistical study

Lars G Johnsen, National Library of Norway

In this talk we report on work that is carried out at the The National Library of Norway[1] in the context of a pilot project aimed at building a thesaurus (see link[2] for a Norwegian description) for improving the inventory of topic words employed in classifications (e.g. Dewey Decimal Classification, DDC).

For this purpose we made a statistic study into the relationship between topic words and content words using the digitized material from the Norwegian National Library. What is the relationship between topic and content words? In general we may expect that topic words are more general than content words[3], they are higher in a semantic taxonomy (e.g. WordNet[4]), or indicate a frame as in FrameNet[5].

A selection of books was chosen to provide the statistical material. Books classified within DDC, approximately 144 000 titles served as input. The classification codes themselves were simplified to the first three digits. These constitute the main classes in the Dewey system. The words within each class were tallied and paired with the code of the book they occurred in, resulting in a data set of triplets consisting of frequency, code and word, like: (17,759,kano). There were no morphological analysis (as in lemmatization), and only single words are counted, meaning that multi word expressions are not taken into account.

The data were analyzed for degree of association between code and content word using Pointwise Mutual Information as measure[6], weighted with the square root of the cooccurence. For example, for Dewey code 641 (food and drink) we get among the top associated words, spices like *pepper* and *salt*, words like *flour*, *milk*, *sugar* and *butter* (translated from Norwegian), measurements like *dl*, *ss* (tablespoon), *g* (grams) among others, all judged to be relevant for the class.

Words that were associated with the classification are then compared to the topic words of that classification, to judge how they relate. The topic words were taken from the electronic catalogues encoded using MARC 21 from approximately 120 000 titles, resulting in a grand total 620 000 occurences of topic words.

For both the topic and content words the grouping from the book is discarded so topic and content is related through Dewey code only. The resulting data set is on the form: frequency, code, MARC tag (650 controlled vocabulary, 653 uncontrolled), topic word. For example, (10, 305, 650, women) is one row of data. The topic words were only subject to sorting on frequency.

A group of classification codes underwent further study to see how they related. We present the details of that in the talk, but the conclusion is that taxonomically the relation goes both ways. A term among the topic words can be both more general and more specific than a content word, while it seems that topic words in general tend to be more oriented towards setting up frames (in the FrameNet sense), as for example for DDC 799 (hunting and fishing) where *fishing* sets up a frame involving a set of activities (*fishing, fly fishing*) with agents (*the fisherman, a salmon, cod* or *trout*), tools (*rods, lines* etc.) and places (*river, lakes*), where words in italic are taken from the content.

The data can be used also to evaluate texts and use the association for (semi) automatic classification of unknown texts.

---

1 http://www.nb.no

2 http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/Tesaurus-forprosjekt

3 Žumer, M., Ed. (2009). National bibliographies in the digital age: guidance and new directions. IFLA Series on Bibliographic Control, vol. 39. München, K.G. Saur.

4 WordNet, . George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38

5 FrameNet: https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf

6 Kenneth Ward Church og Patrick Hanks (1990). "Word association norms, mutual information, and lexicography". Computational Linguistics 16 (1): 22–29.

Abstract Scania (DanTerm 2015)

Text – Ontology – Text

A return journey by truck

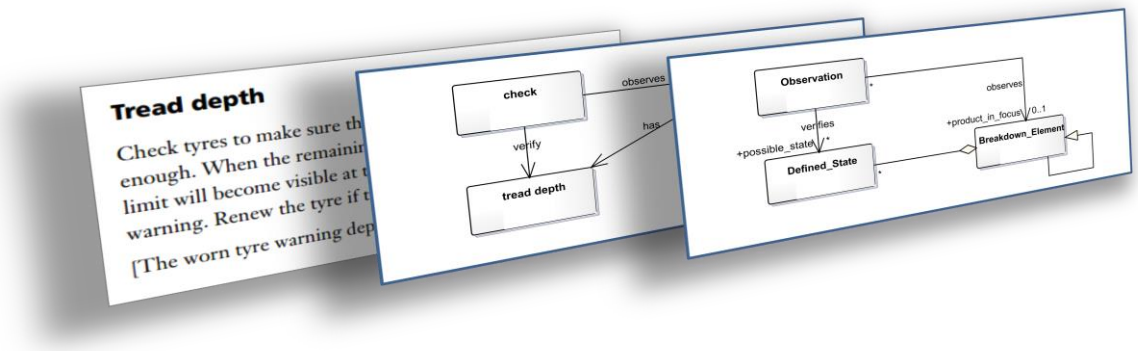
Every vehicle made by Scania has its own specification, described by thousands of parameters that make an almost infinite number of unique combinations. Thus, writing a description that is valid for all vehicles of a certain model is not an option, since each individual vehicle is, so to say, a model of its own.

The key to describing a modular product is modular information. We believe that in the future we will build structures of objects and relations, rather than write text, in order to describe repair and maintenance.

We have made a pilot study where we manually transformed text into structures, by analysing texts and concepts. In these structures we found patterns regarding what kind of objects the structures are built from and how the objects are related to each other. Our term bank Scania Lexicon helped us categorise the different objects.

From the result, we created a model which is now being implemented in our central system for product configuration management.

In this presentation I will show how we worked from text analysis to a model, How we are now using the model for building structures, and how we in the next step will re-generate texts from the structures.



Reference model

ISO 10303-239 (Product Life Cycle Support http://www.plcs.org/ap239/)

From text to ontology

The DanTermBank Project Team, IBC, Copenhagen Business School

The work described here is carried out as part of the DanTermBank project, www.dantermbank.dk. The long-term goal of this project is to establish the foundations for a national terminology and knowledge bank, and consequently the project must find solutions to a wide range of problems, e.g. how to increase the number of terms whilst at the same time maintaining a high quality of information, how to choose, organize and present data suitable for different user groups and how to ensure interoperability with other term banks.

The primary goal of the first phase of the DanTermBank project is to develop automatic methods and prototypes for advanced terminology work. The overall architecture of the DanTermBank project is composed by three main parts:

1. Knowledge acquisition
2. Knowledge validation
3. Knowledge dissemination

The primary aim of subproject 1 is to develop new advanced models of and methods for automatic extraction of concepts and information about concepts from text. The models and methods should, in principle, be applicable to a large number of languages, but for the moment we primarily focus on developing tools that work on Danish language texts.

In order to fulfill this aim, the project is developing a prototype for corpus compilation and processing, which, on the basis of domain texts collected from the Internet, can automatically extract terms and relations and produce a draft version of a terminological ontology. The subproject develops two prototypes (see e.g. Lassen, 2012, pp.218-230): a corpus compilation tool (dtCrawler) and a corpus processing tool (dtbTOOLS). The first tool allows us to compile domain corpora from the web, combining a selection of statistical measures with known terminology input and set operations. The second tool allows us to tag the corpora, and extract knowledge in the form of term candidates as well as relations between them from the corpora, using a combination of POS-patterns, rules, statistical measures, and a combined ranking score.

In our presentation we will demonstrate the prototypes for knowledge acquisition.

References:

Lassen, Tine (2012). A Corpus Compilation and Processing Prototype for Terminology Work. In: Aguado de Cea et al. (Eds.): *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012),* pp.218-230. 19-22 June 2012, Madrid, Spain

Knowledge validation

The DanTermBank Project Team, IBC, Copenhagen Business School

The work described here is carried out as part of the DanTermBank project, www.dantermbank.dk. The long-term goal of this project is to establish the foundations for a national terminology and knowledge bank, and consequently the project must find solutions to a wide range of problems, e.g. how to increase the number of terms whilst at the same time maintaining a high quality of information, how to choose, organize and present data suitable for different user groups and how to ensure interoperability with other term banks.

The primary goal of the first phase of the DanTermBank project is to develop automatic methods and prototypes for advanced terminology work. The overall architecture of the DanTermBank project is composed by three main parts:

1. Knowledge acquisition
2. Knowledge validation
3. Knowledge dissemination

The primary aim of subproject 2 is to develop methods and a prototype that may be used for automatic validation and dynamic expansion of the draft ontologies that result from the automatic knowledge extraction.

Our terminological ontologies contain formalized descriptions of the concepts in the form of typed feature structures. A typed feature structure consists of a number of attribute-value pairs, or feature specifications, each representing a characteristic of a concept. The distribution of feature specifications in a terminological ontology is subject to a number of constraints primarily based on inheritance, and the validation of draft ontologies is based on these constraints.

In our presentation we will give examples of knowledge validation.

References:

Lassen, Tine, Bodil Nistrup Madsen & Hanne Erdman Thomsen (2011). 'Automatic Knowledge Extraction and Knowledge Structuring for a National Term Bank'. In: *NODALIDA 2011 workshop: Creation, Harmonization and Application of Terminology Resources* .NEALT Proceedings Series Vol. 12. http://hdl.handle.net/10062/17274 (2011-05-09), pp 23-26.

OOPS! (OntOlogy Pitfall Scanner!): evaluating ontologies online

The 1990s and the first years of this new millennium have witnessed the growing interest of many practitioners in methodologies that support the creation of ontologies. All these approaches have supposed a step forward since they have transformed the art of building ontologies into an engineering activity. The most complete and well-known methodologies are METHONTOLOGY [1], On-To-Knowledge [4], DILIGENT [3], or the NeOn Methodology [5], which proposes a new approach to developing ontologies based on the reuse of as many knowledge resources as possible. The correct application of such methodologies benefits the ontology quality. However, such a quality is not totally guaranteed because developers must tackle a wide range of difficulties and handicaps when modelling ontologies. These difficulties can imply the appearance of anomalies or errors in ontologies. Therefore, it is important to evaluate the ontologies before using or reusing them in other ontologies and/or semantic applications.

One approach for evaluating ontologies is to analyse whether the ontology is conform to ontology modelling best practices. In other words, to check whether the ontologies contain anomalies or pitfalls.

In this context, our goal is to present an **on-line tool** that supports the **automatic detection of pitfalls in ontologies**. This tool is called OOPS! (OntOlogy Pitfall Scanner! available at http://oops.linkeddata.es) and represents a new option for ontology developers within the range of ontology evaluation tools available. OOPS! **extends the list of pitfalls detected** by most recent and available approaches (like MoKi [2] and XD-Analyzer[1]) and allows **selecting subset of pitfalls** to be analysed according to different evaluation dimensions. As not all the pitfalls are equally important, an **indicator (critical, important, minor)** has been associated to each pitfall according to their possible negative consequences. OOPS! provides **conformance badges** attached to the evaluation results, according to the higher degree of importance (minor, important or) among the detected pitfalls or “free of pitfall” otherwise. These badges could be included within the ontology documentation. In addition, OOPS! is **executed independently** of the ontology development platform without configuration or installation, it works with main web browsers (Firefox, Chrome, Safari and Internet Explorer) and a **RESTFul web service** is also provided. Currently it has been used by developers from more than **48 countries** (>2000 times), embedded within **3 third-party software** developments and used in several **enterprises**, supporting both ontology development processes and training activities.

References

1. Fernández-López, M., Gómez-Pérez, Sierra, J. P., & Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment. *Intelligent Systems and their Applications, IEEE*, *14*(1), 37-46.
2. Pammer, V. *PhD Thesis: Automatic Support for Ontology Evaluation Review of Entailed Statements and Assertional Effects for OWL Ontologies*. Engineering Sciences. Graz University of Technology. 2010.
3. Pinto, H.S., Tempich, C., Staab, S. *DILIGENT: Towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies*. In Ramón López de Mantaras and Lorenza Saitta, Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004), pp. 393--397. IOS Press, Valencia, Spain, August 2004. ISBN: 1-58603-452-9. ISSN: 0922-6389.
4. Staab, S., Schnurr, H.P., Studer, R., Sure, Y. *Knowledge Processes and Ontologies*. IEEE Intelligent Systems 16(1):26–34. (2001).
5. Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E., & Gangemi, A. (Eds.). (2012). *Ontology engineering in a networked world*. Springer.

[1] http://neon-toolkit.org/wiki/XDTools

Logical Characterization of Ontology Construction Using Fuzzy Description Logics

Farshad Badie, Hans Götzsche
{badie-goetzsche@hum.aau.dk}

Center for Linguistics, Department of Communication, Aalborg University

Abstract

Ontologies based on Description Logics (DLs) have proved to be effective in formally sharing knowledge across semantic technologies, e.g. Semantic Web, Natural Language Processing, Text Analytics, Business intelligence. Our main goal is analyzing ontology construction considering vagueness. We have had the extension of ontologies with Fuzzy Logic capabilities which plan to make proper backgrounds for ontology driven reasoning and argumentation on vague and imprecise domains.

This presentation conceptualizes learning from fuzzy classes using the Inductive Logic Programming framework. Then, employs Description Logics in characterizing and analyzing fuzzy statements. And finally, provides a conceptual framework describing fuzzy concept learning in ontologies using the Inductive Logic Programming.

Keywords: Ontology Construction; Knowledge Representation; Fuzzy Description Logic; Inductive Logic Programming; Concept Learning.

Logical Characterization of Ontology Construction Using Fuzzy Description Logics

Farshad Badie, Hans Götzsche
{badie-goetzsche@hum.aau.dk}

Center for Linguistics, Department of Communication, Aalborg University

November 16, 2014

References

[1] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[2] F. Bobillo and U. Straccia. fuzzyDL: An expressive fuzzy description logic reasoner. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Comp. Int.). IEEE Int. Conf. on*, pages 923 –930, june 2008.

[3] F. Bobillo and U. Straccia. Representing fuzzy ontologies in OWL 2. In *Fuzzy Systems (FUZZ), 2010 IEEE Int. Conf. on*, july 2010.

[4] Fernando Bobillo and Umberto Straccia. fuzzyDL: An expressive fuzzy description logic reasoner. In *IEEE International Conference on Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence)*, pages 923–930, 2008.

[5] Fernando Bobillo and Umberto Straccia. Fuzzy ontology representation using owl 2. *CoRR*, abs/1009.3391, 2010.

[6] Silvia Calegari and Davide Ciucci. Fuzzy ontology, fuzzy description logics and fuzzy-owl. In Francesco Masulli, Sushmita Mitra, and Gabriella Pasi, editors, *WILF*, volume 4578 of *Lecture Notes in Computer Science*, pages 118–126. Springer, 2007.

[7] Silvia Calegari and Davide Ciucci. Fuzzy ontology, fuzzy description logics and fuzzy-owl. In Francesco Masulli, Sushmita Mitra, and Gabriella Pasi, editors, *WILF*, volume 4578 of *Lecture Notes in Computer Science*, pages 118–126. Springer, 2007.

[8] Jianhua Chen and Sukhamay Kundu. A sound and complete fuzzy logic system using zadeh's implication operator. In Zbigniew W. Ras and Maciej Michalewicz, editors, *ISMIS*, volume 1079 of *Lecture Notes in Computer Science*, pages 233–242. Springer, 1996.

[9] Xiaochun Cheng, Yunfei Jiang, and Xuhua Liu. The rationality and decidability of fuzzy implications. In *IJCAI*, pages 1910–1915. Morgan Kaufmann, 1995.

[10] Didier Dubois and Henri Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.

[11] Noah D. Goodman, Joshua B. Tenenbaum, Jacob Feldman, and Thomas L. Griffiths. A rational analysis of rule-based concept learning. *Cognitive Science*, 32(1):108–154, 2008.

[12] Josué Iglesias and Jens Lehmann. Towards integrating fuzzy logic capabilities into an ontology-based inductive logic programming framework. In *Proc. of the 11th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2011.

[13] George J. Klir and Bo Yuan. *Fuzzy sets and fuzzy logic - theory and applications.* Prentice Hall, 1995.

[14] Stasinos Konstantopoulos and Angelos Charalambidis. Formulating description logic learning as an inductive logic programming task. In *FUZZ-IEEE*, pages 1–7. IEEE, 2010.

[15] Stasinos Konstantopoulos and Angelos Charalambidis. Formulating description logic learning as an inductive logic programming task. In *Proc. of FUZZ-IEEE, 2010 IEEE World Congress on Comp. Int., July 18–23, Barcelona*. IEEE, July 2010.

[16] Richard C. T. Lee. Fuzzy logic and the resolution principle. *J. ACM*, 19(1):109–119, 1972.

[17] Jens Lehmann. DL-Learner: learning concepts in description logics. *Journal of Machine Learning Research (JMLR)*, 10:2639–2642, 2009.

[18] Jens Lehmann, Sören Auer, Lorenz Bühmann, and Sebastian Tramp. Class expression learning for ontology engineering. *Journal of Web Semantics*, 9:71 – 81, 2011.

[19] Jens Lehmann and Pascal Hitzler. Concept learning in description logics using refinement operators. *Machine Learning journal*, 78(1-2), 2010.

[20] Francesca A. Lisi and Umberto Straccia. An inductive logic programming approach to learning inclusion axioms in fuzzy description logics. In Fabio Fioravanti, editor, *CILC*, volume 810 of *CEUR Workshop Proceedings*, pages 57–71. CEUR-WS.org, 2011.

[21] Chang Liu, Guilin Qi, Haofen Wang, and Yong Yu. Fuzzy reasoning over rdf data using owl vocabulary. In Olivier Boissier, Boualem Benatallah, Mike P. Papazoglou, Zbigniew W. Ras, and Mohand-Said Hacid, editors, *Web Intelligence*, pages 162–169. IEEE Computer Society, 2011.

[22] Alexander Maedche. *Ontology Learning for the Semantic Web*. 2002.

[23] Stephen Muggleton. Inductive logic programming. *New Generation Computing*, 8(4):295–318, 1991.

[24] R. Navigli, P. Velardi, and A. Gangemi. Ontology learning and its application to automated terminology translation. *IEEE Intelligent Systems*, pages 22–31, 2003.

[25] Shan-Hwei Nienhuys-Cheng and Ronald de Wolf. *Foundations of Inductive Logic Programming*, volume 1228 of *LNAI*. Springer-Verlag, 1997.

[26] Guilin Qi, Jeff Z. Pan, and Qiu Ji. A possibilistic extension of description logics. In *Description Logics*, pages 435–442. CEUR-WS.org.

[27] Sebastian Rudolph. Foundations of description logics. In *In Reasoning Web, volume 6848 of LNCS*. Springer, 2011.

[28] Paul J. Silvia. *Exploring the psychology of interest*. Oxford University Press, New York, 2006.

[29] Giorgos Stoilos, Giorgos B. Stamou, and Jeff Z. Pan. Fuzzy extensions of owl: Logical properties and reduction to fuzzy description logics. *Int. J. Approx. Reasoning*, 51(6):656–679.

[30] Umberto Straccia. A Fuzzy Description Logic for the Semantic Web. In *Fuzzy logic and the Semantic Web, Capturing Int., Chapter 4*, pages 167–181. Elsevier, 2005.

[31] Lofti A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

Ontology in disasters

Khurshid Ahmad

Trinity College Dublin

During a disaster, natural or man-made, is characterised by many attributes. However, one attribute that is common to almost all disasters – the need for information, information needed for survival, recovery and rehabilitation, during and after a disaster is in inverse proportion to the availability of such information. The question we ask in this paper is how to organise such information? This sort of ontological question is asked usually for well-grounded subjects in science, technology, and arts and humanities, where plenty of information is available when needed. We look at the case of natural disasters and see how information flows across domains, amongst first responders, state agencies, civic organisations and the disaster impacted. Terms are used frequently during a disaster to avoid ambiguity but the burden of interpretation is usually on the recipient and this reduces the amount of information imparted. We describe our work in progress in Project Slandail, an EU sponsored FP7 Project, where we are attempting to harness and amplify information in different modalities –text, images- released by different agencies, through an ontologically grounded term sets, through an information articulation strategy that emphasises clarity in communications.