ANITA: A Narrative Interpretation of Taxonomies for their Adaptation to Text Collections

Mario Cataldi
Università di Torino
Torino, Italy
cataldi@di.unito.it

K. Selçuk Candan
Arizona State University
Tempe, AZ 85283, USA
candan@asu.edu

Maria Luisa Sapino
Università di Torino
Torino, Italy
mlsapino@di.unito.it

ABSTRACT

[illegible] *A Narrative Interpretation of Taxonomies for their Adaptation* (ANITA) [illegible]

Categories and Subject Descriptors

[illegible] **Hypertext/Hypermedia** [illegible] **Document and Text Processing** [illegible] **Information Search and Retrieval** [illegible]

General Terms

[illegible]

Keywords

[illegible]

1. INTRODUCTION

[illegible] *within the context of* [illegible] *A Narrative Interpretation for Taxonomy Adaptation* (ANITA) [illegible]

2. RELATED WORK

[illegible]

CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.

3. NARRATIVE-DRIVEN TAXONOMY ADAPTATION PROCESS

... $H(C, E)$ ... $C = \{c_1, \ldots, c_n\}$ ... E ... $H'(C', E')$ ... D ... ANITA ...

3.1 Step I: Narrative View of a Taxonomy

$$s\vec{v}_{c_i} = \{w_{i,1}, w_{i,2}, w_{i,3} \cdots w_{i,v}\}$$

3.2 Step II: Segmentation of the Narrative

$$\Delta_{i,i+1} = 1 - cos(\vec{sv}_i, \vec{sv}_{i+1})$$

3.3 Step III: Taxonomy Reconstruction

3.3.1 Step IIIa: Partition Linking

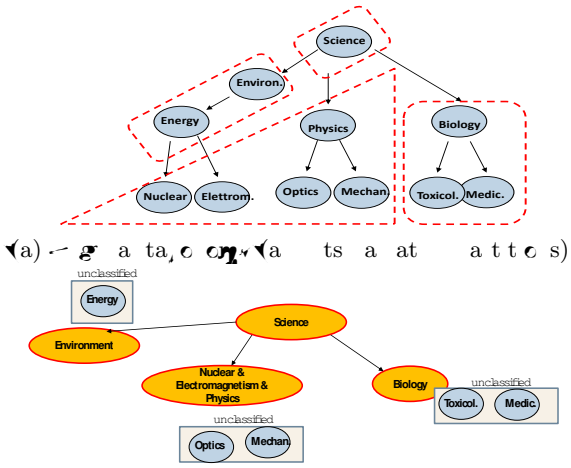

Figure 1: Narrative-based adaptation of a taxonomy fragment: based on the structural constraints and the available NSF content (described in Section4) (a) the partitions are linked to each other. Finally, (b) each partition is labeled by selecting a representative label.

3.3.2 Step IIIb: Partition Labeling

4. USER STUDY

Context: NSF Corpus		
	avg time (sec)	avg num. of interactions
Original (7 concepts)	35	5
ANITA (3 concepts)	7	3
k-Means (3 concepts)		

Table 1: User Study: Average time and average number of interactions (clicks on the structure for expanding or collapsing nodes) per taxonomy, when the users explore the structure to retrieve documents related to a randomly selected concept.

Context: NSF Corpus		
	easy to use	sufficiently detailed
Original (7 concepts)	4	3.8
ANITA (3 concepts)	4	3.6
k-Means (3 concepts)	3.3	6

Table 2: Subjective questions in the user study: for each question, each user has quantified her opinion by a 5-point scale rating.

5. CONCLUSIONS

6. REFERENCES

1. C. Brewster, F. Ciravegna, and Y. Wilks. User-centred ontology learning for knowledge management. In *NLDB '02*, pages 203–207. Springer-Verlag, 2002.
2. P. Buitelaar, P. Cimiano, and B. Magnini, editors. *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press, 2005.
3. M. Cataldi, C. Schifanella, K. S. Candan, M. L. Sapino, and L. Di Caro. Cosena: a context-based search and navigation system. In *MEDES '09*, pages 160–168. ACM, 2009.
4. P. Cimiano, A. Hotho, and S. Staab. Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research*, 24:305–339, 2005.
5. J. W. Kim and K. S. Candan. Cp-cv: concept similarity mining without frequency information from domain describing taxonomies. In *CIKM '06*, 2006.
6. D. J. Lawrie and W. B. Croft. Generating hierarchical summaries for web searches. In *SIGIR'03*, pages 457–458, 2003.
7. C. Muller, I. Gurevych, and M. Muhlhauser. Integrating semantic knowledge into text similarity and information retrieval. In *ICSC '07*, pages 257–264, 2007.
8. S. P. Ponzetto and M. Strube. Deriving a large scale taxonomy from wikipedia. In *AAAI*, pages 1440–1445, 2007.
9. M. Sanderson. Word sense disambiguation and information retrieval. In *SIGIR'94*, pages 142–151. Springer-Verlag New York, Inc., 1994.
10. M. Sanderson and B. Croft. Deriving concept hierarchies from text. In *SIGIR99*, pages 206–213, 1999.
11. O. Zamir and O. Etzioni. Web document clustering: a feasibility demonstration. In *SIGIR'98*, pages 46–54, 1998.
12. Y. Zhao and G. Karypis. Evaluation of hierarchical clustering algorithms for document datasets. In *Data Mining and Knowledge Discovery*, pages 515–524. ACM Press, 2002.