

Effective alignment of search results with content annotation and query value

K.SRINU

JDIT, Yavatmal, India

Abstract

Numerous companies now produce and disseminate product descriptions in written form. The organized information in such textual collections is obscured by the abundance of unstructured material. The extraction of structured relationships is made easier by information extraction techniques, however these methods may be time-consuming, costly, and erroneous. In this paper, we describe a unique alternative method for generating structured metadata, one that does so by locating documents most likely to contain information of interest and which may be used in future database queries.

Search Terms: Annotated documents, dynamic forms, and groupware

1. Introduction

Users generate and disseminate content in a wide variety of application disciplines. The "pay-as-you-go" querying approach is a relatively new area of research that aims to use more expressive queries by making use of such annotations. In order to make it easier to annotate data in the field, we offer a technology called Collaborative Adaptive Data Sharing (CADS). Our system's main innovation is that, in addition to analyzing the document's content, it also uses the query burden to guide the annotating process..

2. Existing system

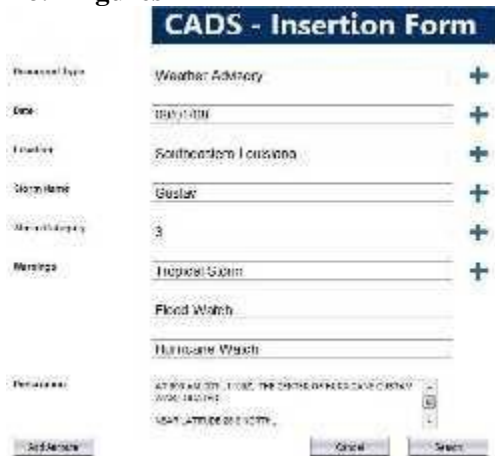
Annotation of "untyped" keywords is sometimes the sole option in annotation systems. The "pay-as-you-go" querying approach makes advantage of the expressiveness of annotations with attribute-value pairs. Because of this, people using data input tend to disregard annotation features.

2.1 Proposed system

We propose CADS (Collaborative Adaptive Data Sharing platform) as an infrastructure for "annotate-as-you-create" annotation of data in the field. To facilitate and reduce the price of producing well-annotated papers is CADS's primary objective. The creator (author) may review the form, make any required changes to the automatically created information, and then save the annotated document.

3. Figures and Tables

3.1 Figures

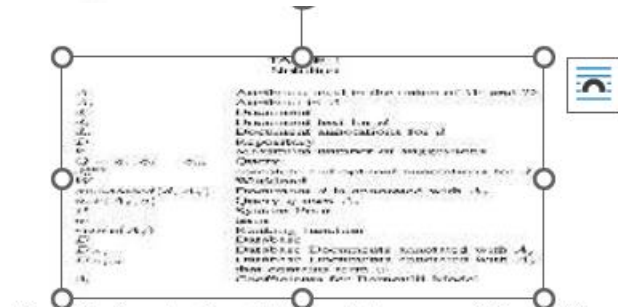


This figure shows the insertion form of CADS.

3.2 Tables

TABLE 1
Attributes

A_1	Attribute used to the value of A_1 and A_2
A_2	Attribute used to the value of A_2
A_3	Attribute used to the value of A_3
A_4	Attribute used to the value of A_4
A_5	Attribute used to the value of A_5
A_6	Attribute used to the value of A_6
A_7	Attribute used to the value of A_7
A_8	Attribute used to the value of A_8
A_9	Attribute used to the value of A_9
A_{10}	Attribute used to the value of A_{10}
A_{11}	Attribute used to the value of A_{11}
A_{12}	Attribute used to the value of A_{12}
A_{13}	Attribute used to the value of A_{13}
A_{14}	Attribute used to the value of A_{14}
A_{15}	Attribute used to the value of A_{15}
A_{16}	Attribute used to the value of A_{16}
A_{17}	Attribute used to the value of A_{17}
A_{18}	Attribute used to the value of A_{18}
A_{19}	Attribute used to the value of A_{19}
A_{20}	Attribute used to the value of A_{20}
A_{21}	Attribute used to the value of A_{21}
A_{22}	Attribute used to the value of A_{22}
A_{23}	Attribute used to the value of A_{23}
A_{24}	Attribute used to the value of A_{24}
A_{25}	Attribute used to the value of A_{25}
A_{26}	Attribute used to the value of A_{26}
A_{27}	Attribute used to the value of A_{27}
A_{28}	Attribute used to the value of A_{28}
A_{29}	Attribute used to the value of A_{29}
A_{30}	Attribute used to the value of A_{30}
A_{31}	Attribute used to the value of A_{31}
A_{32}	Attribute used to the value of A_{32}
A_{33}	Attribute used to the value of A_{33}
A_{34}	Attribute used to the value of A_{34}
A_{35}	Attribute used to the value of A_{35}
A_{36}	Attribute used to the value of A_{36}
A_{37}	Attribute used to the value of A_{37}
A_{38}	Attribute used to the value of A_{38}
A_{39}	Attribute used to the value of A_{39}
A_{40}	Attribute used to the value of A_{40}
A_{41}	Attribute used to the value of A_{41}
A_{42}	Attribute used to the value of A_{42}
A_{43}	Attribute used to the value of A_{43}
A_{44}	Attribute used to the value of A_{44}
A_{45}	Attribute used to the value of A_{45}
A_{46}	Attribute used to the value of A_{46}
A_{47}	Attribute used to the value of A_{47}
A_{48}	Attribute used to the value of A_{48}
A_{49}	Attribute used to the value of A_{49}
A_{50}	Attribute used to the value of A_{50}
A_{51}	Attribute used to the value of A_{51}
A_{52}	Attribute used to the value of A_{52}
A_{53}	Attribute used to the value of A_{53}
A_{54}	Attribute used to the value of A_{54}
A_{55}	Attribute used to the value of A_{55}
A_{56}	Attribute used to the value of A_{56}
A_{57}	Attribute used to the value of A_{57}
A_{58}	Attribute used to the value of A_{58}
A_{59}	Attribute used to the value of A_{59}
A_{60}	Attribute used to the value of A_{60}
A_{61}	Attribute used to the value of A_{61}
A_{62}	Attribute used to the value of A_{62}
A_{63}	Attribute used to the value of A_{63}
A_{64}	Attribute used to the value of A_{64}
A_{65}	Attribute used to the value of A_{65}
A_{66}	Attribute used to the value of A_{66}
A_{67}	Attribute used to the value of A_{67}
A_{68}	Attribute used to the value of A_{68}
A_{69}	Attribute used to the value of A_{69}
A_{70}	Attribute used to the value of A_{70}
A_{71}	Attribute used to the value of A_{71}
A_{72}	Attribute used to the value of A_{72}
A_{73}	Attribute used to the value of A_{73}
A_{74}	Attribute used to the value of A_{74}
A_{75}	Attribute used to the value of A_{75}
A_{76}	Attribute used to the value of A_{76}
A_{77}	Attribute used to the value of A_{77}
A_{78}	Attribute used to the value of A_{78}
A_{79}	Attribute used to the value of A_{79}
A_{80}	Attribute used to the value of A_{80}
A_{81}	Attribute used to the value of A_{81}
A_{82}	Attribute used to the value of A_{82}
A_{83}	Attribute used to the value of A_{83}
A_{84}	Attribute used to the value of A_{84}
A_{85}	Attribute used to the value of A_{85}
A_{86}	Attribute used to the value of A_{86}
A_{87}	Attribute used to the value of A_{87}
A_{88}	Attribute used to the value of A_{88}
A_{89}	Attribute used to the value of A_{89}
A_{90}	Attribute used to the value of A_{90}
A_{91}	Attribute used to the value of A_{91}
A_{92}	Attribute used to the value of A_{92}
A_{93}	Attribute used to the value of A_{93}
A_{94}	Attribute used to the value of A_{94}
A_{95}	Attribute used to the value of A_{95}
A_{96}	Attribute used to the value of A_{96}
A_{97}	Attribute used to the value of A_{97}
A_{98}	Attribute used to the value of A_{98}
A_{99}	Attribute used to the value of A_{99}
A_{100}	Attribute used to the value of A_{100}



this table denotes the attributes that are used for notation.

TABLE 2
 Corpus Statistics

	Emergency	CNET	Amazon
Number of Documents	270	4540	19700
Maximum Size (KB)	35.8	504	40
Average Size (KB)	2.87	10.56	1.95
Minimum Size (KB)	0.37	0.16	0.002
Annotations per Document (Max)	24	75	43
Annotations per Document (Min)	1	1	1
Annotations per Document (Avg)	7.9	9.61	4.82

This table denotes the corpus statistics of the document.

4. conclusion

We presented two methods for combining these two types of evidence—content value and querying value—to propose adaptive techniques to suggest relevant attributes to annotate a document. Using the query burden, we demonstrate how the annotation process and the value of collective data may be considerably enhanced.

Acknowledgments

The research of Vagelis Hristidis was funded in part by NSF grants IIS-1216032 and IIS-1216007. Kellner Faculty sponsored by National Science Foundation grant IIS-0643846, and a George A. Kellner Faculty Award recipient, Panagiotis G. Ipeirotis contributed significantly to this work.

References

[1] B. Russell, A. Torralba, K. Murphy, and W. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," *International Journal of Computer Vision*, volume 77, issue 2, pages 157–173, 2008, 1126300700908.

According to [2] "A Quality-Aware Optimizer for Information Extraction," published in *ACM Trans. Database Systems*, volume 34, article 5 in 2009, A. Jain and P.G. Ipeirotis.

For further information on optimal aggregation algorithms for middleware, see [3] R. Fagin, A. Lotem, and M. Naor's article from the June 2003 issue of the *Journal of Computer Systems Sciences* (<http://portal.acm.org/citation.cfm?id=861182.861185>).

As cited in "Tag Ranking," by D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang in *Proceedings of the 18th International Conference on the World Wide Web (WWW)*, 2009.