

EFFICIENT SEARCH RESULT ALIGNMENT WITH ANNOTATION OF CONTENT AND QUERYING VALUE

A.krishna priya, S. Rajiya sulthana.

CSE, BCETW, kadapa, AP, India.
CSE, BCETW, kadapa, AP, India

Abstract

A large number of organizations today generate and share textual descriptions of their products. Such collections of textual data contain some amount of structured information, which remains buried in the unstructured text. While information extraction algorithms facilitate the extraction of structured relations, and are often expensive and inaccurate. we present a novel alternative approach that facilitates the generation of the structured metadata by identifying documents that are likely to contain information of interest and this information is going to be subsequently useful for querying the database.

Keywords - Document annotation, adaptive forms, collaborative platforms

1.Introduction

There are many application domains where users create and share information . A recent line of work toward using more expressive queries that leverage such annotations, is the “pay-as-you-go” querying strategy. we propose Collaborative Adaptive Data Sharing platform (CADS), which is an “annotate-as-you-create” infrastructure that facilitates fielded data annotation. A key contribution of our system is the direct use of the query workload to direct the annotation process, in addition to examining the content of the document.

2. Existing system

Many annotation systems allow only “untyped” keyword annotation: Annotation strategies that use attribute-value pairs are generally more expressive, such annotations, is the “pay- as-you-go” querying

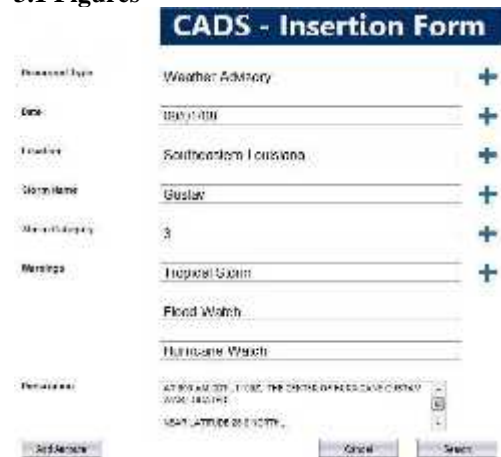
strategy. This results in data entry users ignoring such annotation capabilities.

2.1 Proposed system

In this we propose CADS (Collaborative Adaptive Data Sharing platform), which is an “annotate-as-you create” infrastructure that facilitates fielded data annotation. The goal of CADS is to encourage and lower the cost of creating nicely annotated documents The author (creator) can inspect the form, modify the generated metadata as- necessary, and submit the annotated document for storage.

3. Figures and Tables

3.1 Figures



This figure shows the insertion form of CADS.

3.2 Tables

TABLE 1
Attributes

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

This table denotes the attributes that are used for notation.

TABLE 2
Corpus Statistics

	Emergency	CNET	Amazon
Number of Documents	270	4440	19700
Maximum Size (KB)	35.8	50.4	4.0
Average Size (KB)	3.87	10.56	1.05
Minimum Size (KB)	0.37	0.16	0.002
Annotations per Document (Max)	24	73	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 proposed adaptive techniques to suggest relevant attributes to annotate a document, We present two ways to combine these two pieces of evidence, content value and querying value: we show that using the query workload can greatly improve the annotation process and increase the utility of shared data.

5. Acknowledgments

Vagelis Hristidis was partially supported by US National Science Foundation (NSF) grants IIS-1216032 and IIS-1216007. Panagiotis G. Ipeirotis was partially Kellner Faculty supported by NSF grant IIS-0643846, and a George A. Kellner Faculty supported by NSF grant IIS-0643846, and a George A.

References

[1] B. Russell, A. Torralba, K. Murphy, and W. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," *Int'l J. Computer Vision*, vol. 77, pp. 157-173, 11263-007-0090-8, 2008, doi: 10.1007/s11263-007-0090-8.

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

[3] R. Fagin, A. Lotem, and M. Naor, "Optimal Aggregation Algorithms for Middleware," *J. Computer Systems Sciences*, vol. 66, pp. 614-656, <http://portal.acm.org/citation.cfm?id=861182.861185>, June 2003.

[4] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang, "Tag Ranking," *Proc. 18th Int'l Conf. World Wide Web (WWW)*, 2009.