Captured at: 12/31/2024 12:21 PM URL: https://ieeexplore.ieee.org/document/331726

Precision and recall of GlOSS estimators for database discovery

Publisher: IEEE Cite This PDF

L. Gravano ; H. Garcia-Molina ; A. Tomasic All Authors

10 Cites in Papers
3 Cites in Patents
67 Full Text Views

Abstract
Authors
References
Citations
Keywords
Metrics

Abstract:
Online information vendors and the Internet together offer thousands of text databases from which a user may choose for a given information need. This paper presents a framework for and analyses a solution to this problem, which we call the text-database discovery problem. Our solution is to build a service that can suggest potentially good databases to search. A user's query goes through two steps: first, the query is presented to the GlOSS server (Glossary-Of-Servers Server) to select a set of promising databases to search. Secondly, the query is actually evaluated in the chosen databases. GlOSS gives a hint of what databases might be useful for the user's query, based on word-frequency information for each database. This information indicates how many documents in each database actually contain a keyword, for each field designator. To evaluate the set of databases that GlOSS returns for a given query, we present a framework based on the precision and recall metrics of information retrieval theory. We define metrics for the text-database discovery problem. We further extend our framework by offering different definitions for a "relevant database". We have performed experiments using query traces from the FOLIO library information retrieval system, involving six databases available through FOLIO. The results obtained for different variants of GlOSS are very promising. Even though GlOSS keeps a small amount of information about the contents of the available databases, this information proved to be sufficient to produce very useful hints on where to search.< >

Published in: Proceedings of 3rd International Conference on Parallel and Distributed Information Systems

Date of Conference: 28-30 September 1994
DOI: 10.1109/PDIS.1994.331726
Date Added to IEEE *Xplore*: 06 August 2002
Publisher: IEEE
Print ISBN:0-8186-6400-2
Conference Location: Austin, TX, USA

First Page of the Article



Hide First Page Preview

Authors
References
Citations
Keywords
Metrics

Need Full-Text access to IEEE *Xplore* for your organization?
CONTACT IEEE TO SUBSCRIBE >

More Like This
Construction of Library Information Retrieval System Based on Data Mining
2022 International Symposium on Computing and Artificial Intelligence (ISCAI)
Published: 2022
Application Analysis of Data Mining technology in Internet Information Retrieval
2023 9th International Conference on Systems and Informatics (ICSAI)
Published: 2023
Show More

Get Published in the *IEEE Open Journal of Instrumentation and Measurement*
Learn More

IEEE Personal Account
CHANGE USERNAME/PASSWORD
Purchase Details
PAYMENT OPTIONS
VIEW PURCHASED DOCUMENTS
Profile Information
COMMUNICATIONS PREFERENCES
PROFESSION AND EDUCATION
TECHNICAL INTERESTS
Need Help?
US & CANADA: +1 800 678 4333
WORLDWIDE: +1 732 981 0060
CONTACT & SUPPORT
Follow

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | IEEE Ethics Reporting | Sitemap | IEEE Privacy Policy
A public charity, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2024 IEEE - All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies.

Feedback

IPR2025-00623
UIPATH EXHIBIT 1021