

Chapter 2

Information Retrieval Models: Trends and Techniques

Saruladha Krishnamurthy
Pondicherry Engineering College, India

Akila V
Pondicherry Engineering College, India

ABSTRACT

Information retrieval is currently an active research field with the evolution of World wide web. The objective of this chapter is to provide an insight into the information retrieval definitions, process, models. Further how traditional information retrieval has evolved and adapted for search engines is also discussed. The information retrieval models have not only been used for search purpose it also supports cross lingual translation and retrieval tasks. This chapter also outlines the CLIR process in a brief manner. The tools which are usually used for experimental and research purpose is also discussed. This chapter is organized as Introduction to the concepts of information retrieval. Description of the information retrieval process, the information retrieval models, the role of external sources like ontologies in information retrieval systems. Finally the chapter provides an overview of CLIR and the tools used in developing IR systems is mentioned. Further the latest research directions in IR is explained.

DOI: 10.4018/978-1-5225-2483-0.ch002

INTRODUCTION

Information retrieval(IR) is a field concerned with structure, analysis, storage, organization searching and retrieval of information[Salton,1968]. With the abundant growth of information of web the information retrieval models proposed for retrieval of text documents from books in early 1960's has gained greater importance and popularity among information retrieval scientist and researchers. Today search engine is driven by these information retrieval models.

The fundamental research of information retrieval system focused on searching and retrieving documents relevant to the user information need expressed in the form of query. The challenge lies in retrieving most relevant documents from large corpus by processing the unstructured query. Eventually the information retrieval systems have been designed and researched for retrieving non textual content like video, images, audio and music.

The major issues in IR research is

- **Relevance:**
 - The relevance refer to the retrieval of the information which could be text, audio, image or video from the information sources as requested by an user. The relevance of the retrieval results is user centric as the perspective of relevance varies from one user to other.Designing information retrieval algorithms to retrieve user relevant documents and achieving better retrieval effectiveness is a real challenge.
- **Expression of User's Information Need:**
 - The expectation of the user posing a query could be to expect the information what he/she had in his/her mind. But the problem lies in whether the user expresses his/her needs correctly and precisely. An exact match of the user query to the document may not fetch the relevant documents. The terms used to express the user need in the form of query may not be present in the vocabulary/thesaurus/knowledge source and in literature this is reported as vocabulary mismatch problem or sparse data problem. Though the query given by an user is expanded using the vocabulary the vocabulary must be updated to reflect the terms, phrases currently practiced/used by the user community.Another reason which reduces relevance is that most of the information retrieval systems ignore linguistic relevance and they fetch documents based on the statistical properties.Hence the design of information retrieval systems should take into consideration the linguistic features and user context to fetch more relevant documents /information even though the user query is expressed with less preciseness.

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/information-retrieval-models/178364

Related Content

A Look back at the PAKDD Data Mining Competition 2006

Nathaniel B. Norieland Chew Lim Tan (2007). *International Journal of Data Warehousing and Mining* (pp. 1-11).

www.irma-international.org/article/look-back-pakdd-data-mining/1780

Agent-Based Modeling for Carpooling

Luk Knapen, Ansar-Ul-Haque Yasar, Sungjin Choand Tom Bellemans (2014). *Data Science and Simulation in Transportation Research* (pp. 232-258).

www.irma-international.org/chapter/agent-based-modeling-for-carpooling/90074

Baseline Drift Removal of ECG Signal: Comparative Analysis of Filtering Techniques

Akash Kumar Bhoi, Karma Sonam Sherpaand Bidita Khandelwal (2016). *Research Advances in the Integration of Big Data and Smart Computing* (pp. 134-152).

www.irma-international.org/chapter/baseline-drift-removal-of-ecg-signal/139400

SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence

Rafael Berlanga, Lisette García-Moya, Victoria Nebot, María José Aramburu, Ismael Sanzand Dolores María Llidó (2015). *International Journal of Data Warehousing and Mining* (pp. 1-28).

www.irma-international.org/article/slod-bi/130664

Construction and Application of a Big Data Analysis Platform for College Music Education for College Students' Mental Health

Xiaochen Wangand Tao Wang (2023). *International Journal of Data Warehousing and Mining* (pp. 1-16).

www.irma-international.org/article/construction-and-application-of-a-big-data-analysis-platform-for-college-music-education-for-college-students-mental-health/324060