Chapter 9
Clear and Private Ad Hoc Retrieval Models on Web Data

Souria Ortiga
Mente Argentina University, Argentina

ABSTRACT

During the 1980s, and despite its maturity, the search information (RI) was only intended for librarians and experts in the field of information. Such tendentious vision prevailed for many years. Since the mid-90s, the web has become an increasingly crucial source of information, which has a renewed interest in IR. In the last decade, the popularization of computers, the terrible explosion in the amount of unstructured data, internal documents, and corporate collections, and the huge and growing number of internet document sources have deeply shaken the relationship between man and information. Today, a great change has taken place, and the RI is often used by billions of people around the world. Simply, the need for automated methods for efficient access to this huge amount of digital information has become more important, and appears as a necessity.

INTRODUCTION

During the 1980s and despite its maturity, the search information (RI) was only intended for librarians and experts in the field of information. Such tendentious vision prevailed for many years. Since the mid 90s, the Web has become a source of information increasingly crucial, which has a renewed interest in IR. In the last decade, the popularization of computers, the terrible explosion in the amount of unstructured data, internal documents and corporate collections, and the huge and growing number of internet document sources, have deeply shaken the relationship

DOI: 10.4018/978-1-5225-7338-8.ch009

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
between man and information. Today, in modern life a great change has taken place and the RI is often used by billions of people around the world. However, we are awash in a rising tide of information that the social web has had a wider impact on all sectors of our life. Simply, the need for automated methods for efficient access to this huge amount of digital information has become more important, and appears as a necessity (Frakes, 1992).

IR is the discipline that deals with the search for unstructured data in response to a need for user, which can itself be unstructured, for example, a sentence, a profile or even another document, which can be as structured, for example, a Boolean expression.

In this chapter we highlight an overview of the state of the art on different spots and IR issues we discussed during this chapter are: The ad hoc research, clear and private; information filtering and especially spam filtering problems and plagiarism detection;

**AD-HOC RESEARCH (AR)**

The ad-hoc research (RA) is the standard task in classical IR, based on the interrogation of the information elements (documents in the collection) by the user to obtain the necessary documents after a specified query. The RA has recently conquered the world, fueling not only sought engines in the web, but also any type of unstructured research behind the great web ecommerce. The objective of this task is to automate the document analysis process calculates the comparison between the representation of the need for the information (query) and representation of documents (Larson, 2010).

RA is a process quite familiar to most of us because we all probably use Google at least once a day on average. This task order is to maintain a collection of documents and when a new request comes, we seek in this collection to identify the appropriate documents (called relevant) for this request. The need for information is supposed to be on time rather than long term (as the case in the filtering task see section) and one request at a time is compared to a static document collection. This type of research provides an open field for the user to specify what he needs as a query without any restrictions. Finally,(Kowalski, 2006).

**General Architecture of an Ad-Hoc Research Model (MRA)**

A research model Ad-Hoc (MRA) is a process that stores and manages information on documents, often text documents but can also be multimedia (pictures or video). For example we have some query, q, which is an expression of user needs. An MRA
Web Mining: Creating Structure out of Chaos
www.igi-global.com/chapter/web-mining-creating-structure-out/25769?camid=4v1a