The Mirror DBMS at TREC-8

Arjen P. de Vries and Djoerd Hiemstra
{arjen,hiemstra}@ctit.utwente.nl
CTIT, University of Twente
The Netherlands

Abstract

The database group at University of Twente participates in TREC-8 using the Mirror DBMS, a prototype database system especially designed for multimedia and web retrieval. From a database perspective, the purpose has been to check whether we can get sufficient performance, and to prepare for the very large corpus track in which we plan to participate next year. From an IR perspective, the experiments have been designed to learn more about the effect of the global statistics on the ranking.

1 Introduction

The Mirror DBMS [dV99] combines content management and data management in a single system. The main advantage of such integration is the facility to combine IR with traditional data retrieval. Furthermore, IR researchers can experiment more easily with new retrieval models, using and combining various sources of information. This is an important benefit for advanced IR research; web retrieval, speech retrieval, and cross-language retrieval, each require the use of several representations of content, which is hard to handle in the traditional file-based approach, and becomes too slow in traditional database systems.

In the Mirror DBMS, the IR retrieval model is completely integrated in the database architecture, emphasizing efficient set-oriented query processing. The support for information retrieval in our system is presented in detail in [dV98] and [dVW99]. It supports other types of media as well, which has been demonstrated in the image retrieval system prototype described in [dVvDBA99]. The main goal of our participation in TREC is to test if our system can handle larger data sets without too many problems. Also, we wanted to find out the effect of global statistics on the ranking.

This paper is organized as follows. Sections 2 and 3 review the design of the Mirror DBMS and its support for IR, and discuss its use for TREC processing. Section 4 explains the experimental setup and interprets our results. Section 5 discusses our experience with using the Mirror DBMS for TREC, followed by conclusions.

2 Design

A complete overview and motivation of all aspects of the design of the Mirror DBMS is presented in [dV99]. Although following a traditional three-schema architecture, it uses different