

File ID 337731  
Filename Part II: Exploiting structured resources: introduction

---

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type Dissertation  
Title Effective focused retrieval by exploiting query context and document structure  
Author A.M. Kaptein  
Faculty Faculty of Humanities  
Faculty of Science  
Year 2011  
Pages ix, 192  
ISBN 978-90-814485-7-4

FULL BIBLIOGRAPHIC DETAILS:

<http://dare.uva.nl/record/395691>

---

*Copyright*

*It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use.*

---

## **Part II**

# **Exploiting Structured Resources**



## **Exploiting Structured Resources**

In the second part of this thesis we study how we can make use of the structured resource Wikipedia to retrieve documents and entities. Using Wikipedia as our knowledge resource, we can take advantage of its encyclopedic structure. We move away from a test collection that is based on an unstructured part of the Web i.e., the .GOV2 collection used in the previous part of this thesis, to a test collection that includes a structured part of the Web, namely Wikipedia. Although we are still facing the same challenges of shallowness on the query side and shallowness on the result side, we can now exploit the opportunities that arise from working with a structured resource.

Continuing the work in the previous part, adding query context, we focus on the use of category information as query context. Category information can be given together with the query as explicit information, or it can be implicitly gathered from the data. Category information is of vital importance to a special type of search, namely entity ranking. Entity ranking is the task of finding documents representing entities of an appropriate entity type or category that are relevant to a query.

Ranking entities on the Web is much more complicated than ranking entities in Wikipedia, because a single entity can have many pages on the Web. Search results will be dominated by the most popular one or two entities on the Web, pushing down other relevant entities. Since Wikipedia is structured as an encyclopedia, each entity occurs in principle only once and we do not have to worry about redundant information.

In Chapter 3 we investigate the retrieval of entities and documents inside Wikipedia, while in Chapter 4 we examine how we can retrieve Web homepages of entities using Wikipedia as a pivot.