

Search in Early New High German

„Eyn wunderbarliche vnd seltzame History / von Dyll Uln- spiegel“

14. Juli 2006

ANDREAS HAUSER

Abstract

Searching in Early New High German presents ...

DRAFT

Inhalt

1	Introduction	1
1.1	The Textbase	1
2	Work	2

1 Introduction

1.1 The Textbase

DRAFT

2 Work

In 2003 the first project, which originated in the German linguistics department of the Ludwig-Maximilian-University Munich, has made available the "Till Eulenspiegel" text, which is one of the bigger works of the Early New High German area.

A minor part of the project was using computer technology to aid the research, so the computational linguistics department became involved.

Since the project focused on the compound structures of language it was soon decided to help the extraction and examination by a search engine with a web interface. A quite extensive list of stopwords helped a full text search engine written in Perl which allowed queries with regular expressions. To help the many researchers without prior regular expression knowledge the web interface had a small number of GUI options, which were shortcuts for regular expressions.

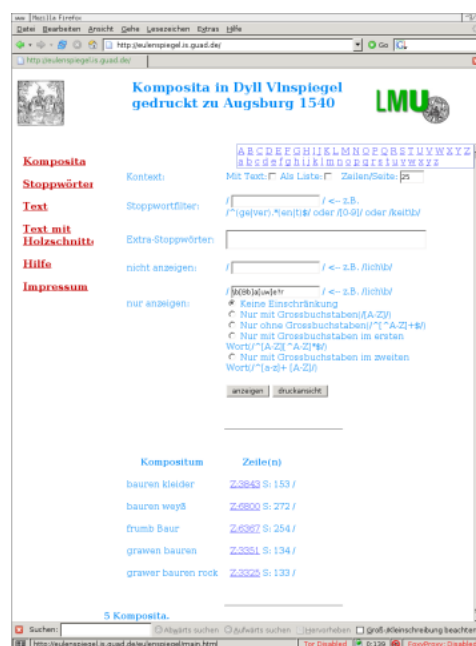


Abbildung 2.1 Screenshot of the Eulenspiegel web interface, 2003

In 2005 work began to extend the search engine with respect to the linguistic aspects, corpora, the Computer Aided aspect, the power of the queries and the web interface. The linguistic

The corpora were extended from the one Eulenspiegel text to 11 texts all from the Early New High German area.

The Computer Aided part evolved from mere finding and managing the data to a more general tool, that helped e.g. part of speech tagging, which was directly done

via the search engines web interface. The speed of the manual work increased and was more fun.

The power of the queries had been limited to compounds before but now was extended to all words and the new information made available like part of speech data or the

The web interface became more dynamic as technologies like JavaScript became more and more accepted and optically more pleasing.

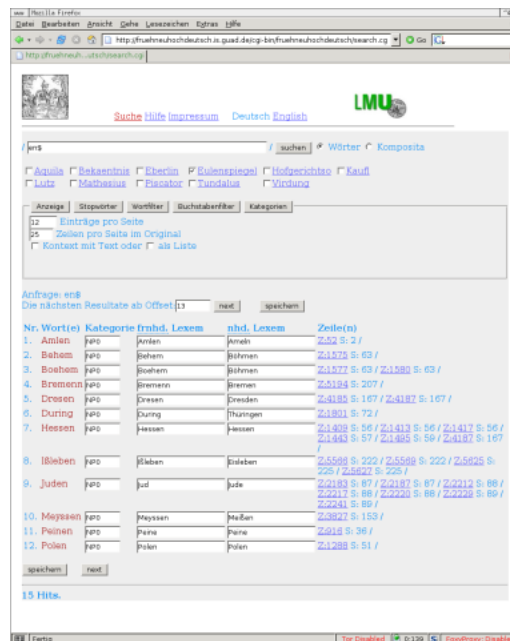


Abbildung 2.2 Screenshot of the evolved web interface, 2005