

Océ@CLEF

**participation of a corporate research
group**



Presentation Outline

- **Océ Mission**
- **Venlo Research Mission**

- **Our participation in 2001**
- **Our participation this year**

- **Evaluation**
- **A critical note**



Océ Mission

“... to enable people to share information by offering products and services for the reproduction, presentation distribution and management of documents.”



Venlo Research Mission

“... to enable Océ to change and to support long term strategy development (on behalf of the continuity of the organization).”



Our participation in 2001

- **Our own ranking model, somewhat like extended boolean models**
- **An unofficial run using “simple proven approaches to text retrieval” (Robertson, Jones)**
- **Turns out to perform a lot better**

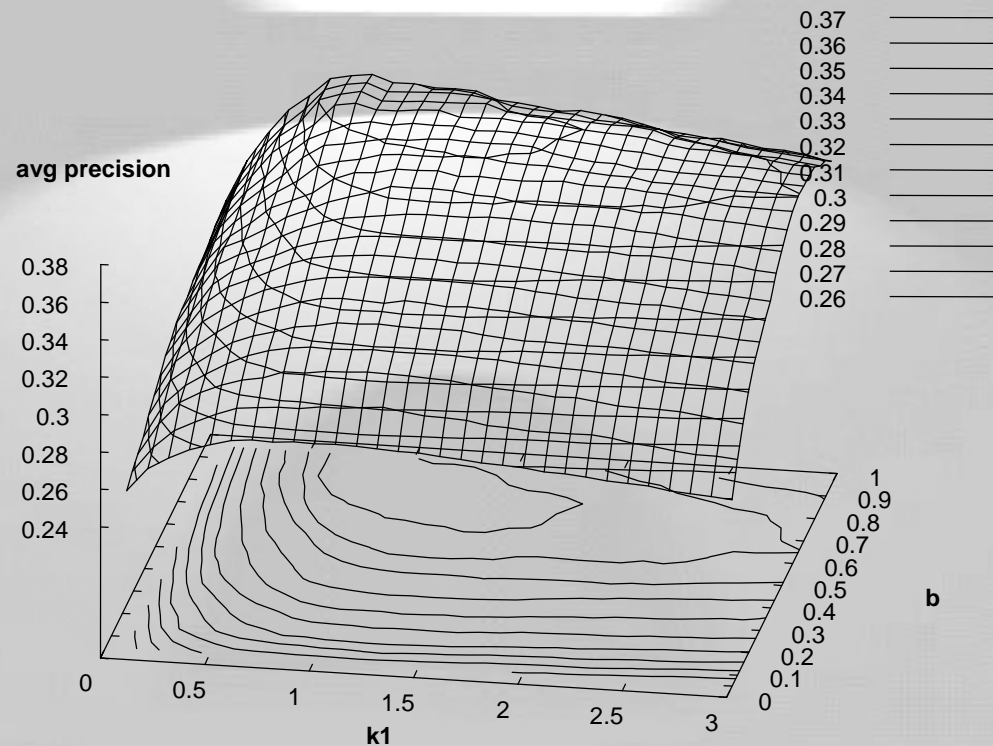


Our participation this year

Ranking

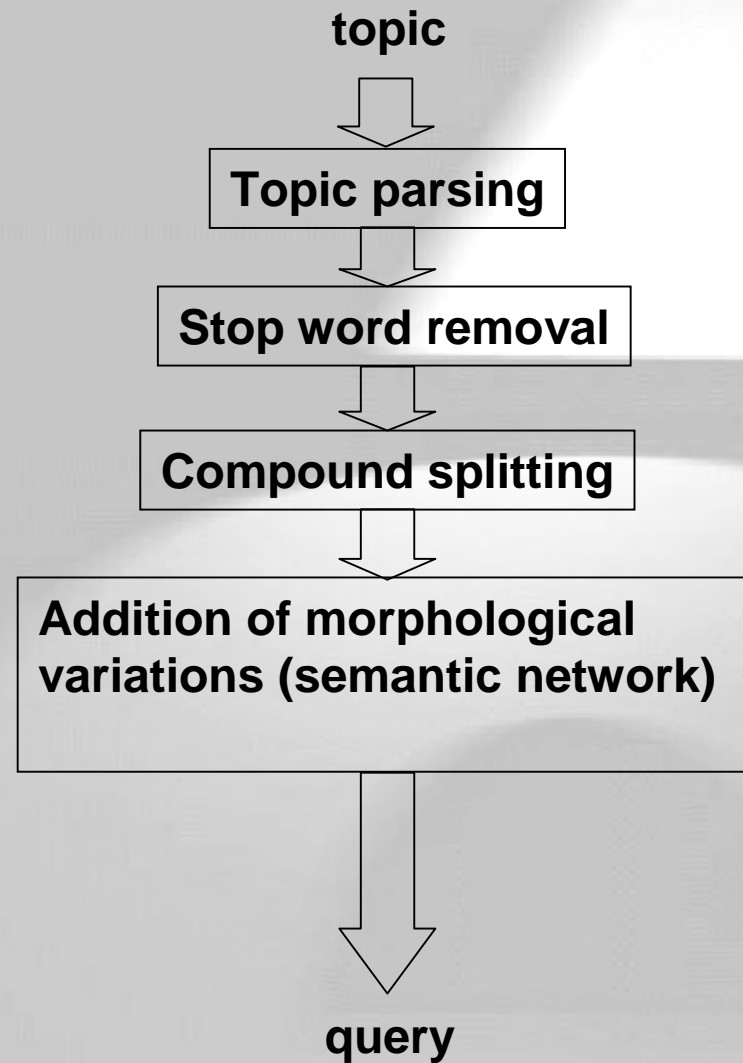
- **BM25**
- **k1 and b determined by brute force parameter space search using 2001 topics**

Dutch: maximum at $k1=0.8, b=0.85$



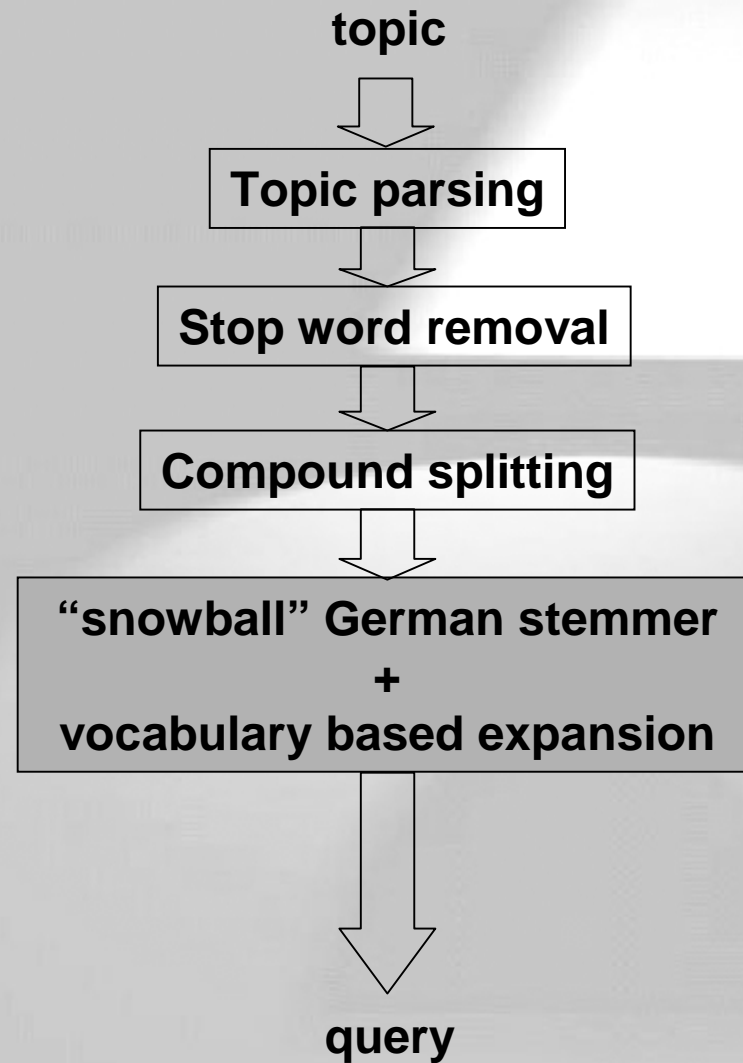
Our participation this year

query construction



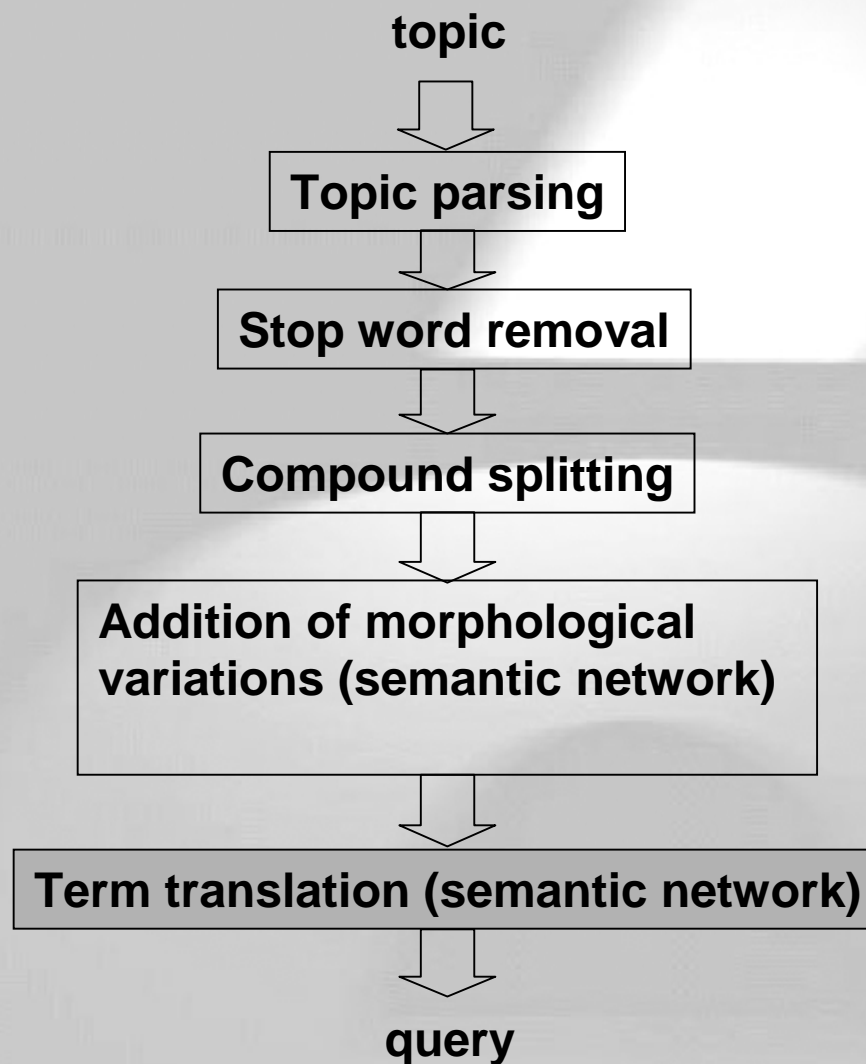
Our participation this year

query construction for German



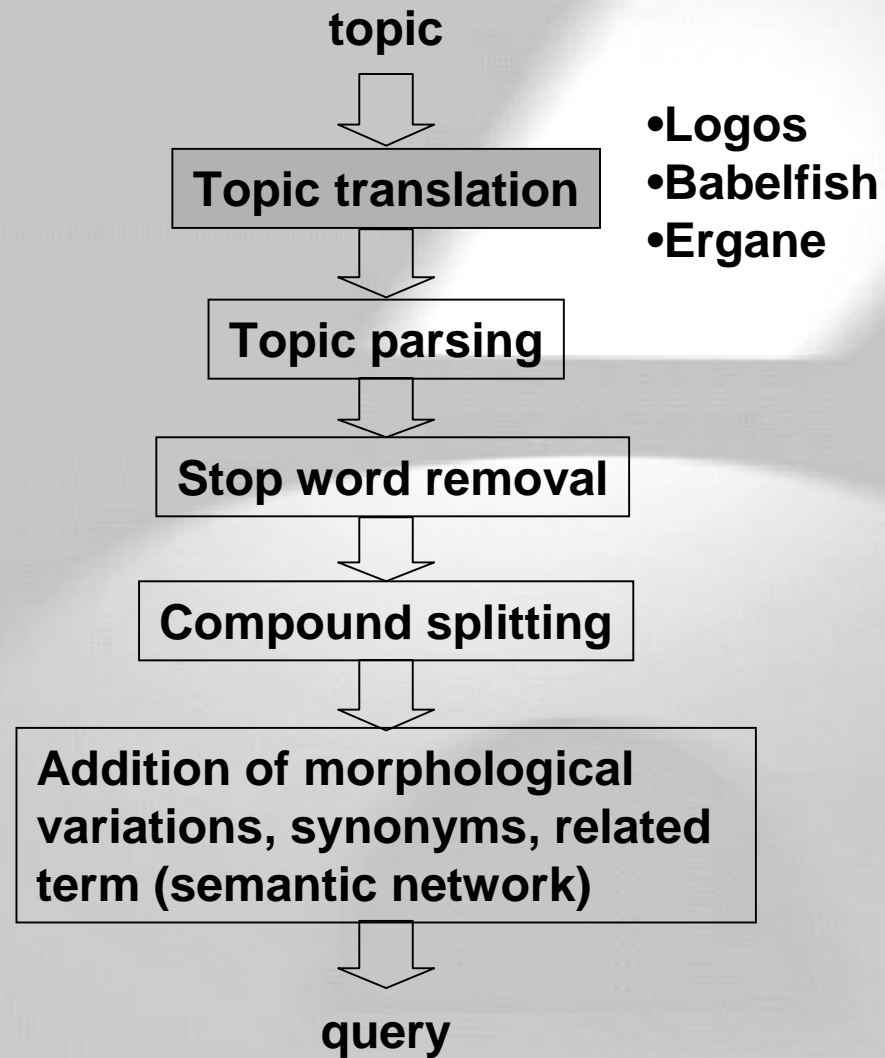
Our participation this year

query construction for cross language tasks



Our participation this year

query construction for cross language tasks (2)



Our participation this year

Result merging in multi lingual task

- **Just the two most basic approaches:**
 - Round Robin
 - Sorted on normalized rank values



Evaluation

- **We used commonly available techniques and algorithms**
- **And scored relatively well**

?

A critical note

- **Results can be attributed to our choice for “easy” languages**
 - (English, Spanish) for bilingual
 - English source for multi lingual
- **It would be better not to compare apples to oranges**
- **Need to limit the number of languages in the contest**

