

Slide 1

**Bilingual Information Retrieval
with HyREX
and Internet Translation Services**

Norbert Gövert

University of Dortmund, Germany
govert@ls6.cs.uni-dortmund.de

Slide 2

Don't be disappointed...

- ... we are beginners in CLIR
- ... we didn't even do 'real' CLIR (according to Peter's claim)
- ... we only reached low precision
- ... we don't take specific measures for domain-specific CLIR

Slide 3

But...

- ... fully automatic approach
- ... cheap approach
- ... IR system design which allows for easy integration of BLIR capabilities
- ... results in terms of recall and precision
- ... starting point for a discussion

Slide 4

Structure of talk

- Data independence in HyREX
- Datatypes and (vague) predicates
- Search predicates for bilingual IR
- Results

Data independence in HyREX

Slide 5

- provide appropriate search operations
- cope with heterogeneous documents
(needed in networked IR, structured document retrieval)
- achieve a modular system design

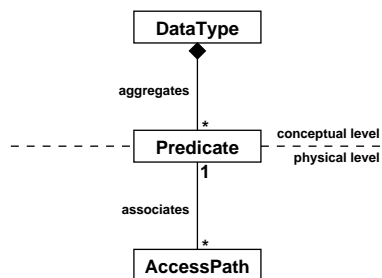
Datatypes and (vague) predicates

$Schema ::= \{AName_1 : Datatype_1, \dots, AName_n : Datatype_n\}$

Example:

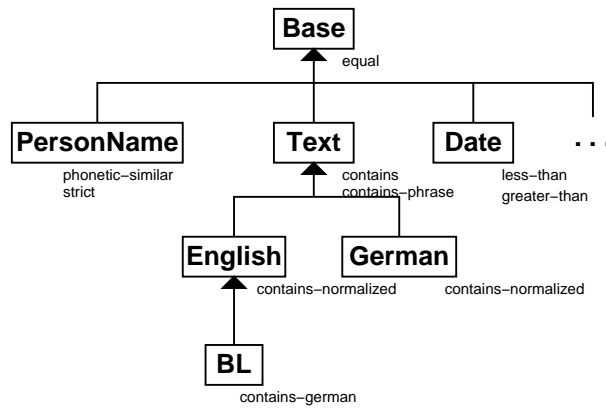
$Schema ::= \{Author : PersonName, Content : Text, PubDate : Date\}$

Slide 6



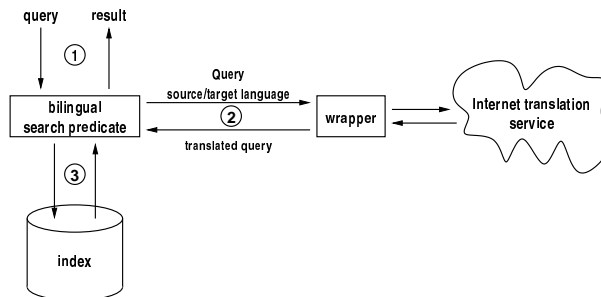
- Author *sounds-like* Norbert Fuhr
- PubDate *around-year* 1999
- Content *contains-phrase* probabilistic IR

Datatypes and (vague) predicates



Slide 7

Search predicates for CLIR



Slide 8

Approach 1 AltaVista's Babelfish (Systran)

Search predicates for CLIR II

Approach 2 Leo's online dictionary (English ↔ German)

Slide 9

LEO German English Dictionary – Access Rate too high

Dear Dictionary user,

Unfortunately we are forced to discontinue this service for you at the moment, because your access rate [...] is too high [...]

Results

Slide 10

