averbis
medical language technology

UNIVERSITÄTS KLINIKUM FREIBURG

Multilingual Access to Biomedical Documents

Stefan Schulz, Philipp Daumke

- Institute of Medical Biometry and Medical Informatics
University Medical Center Freiburg, Germany
- Averbis GmbH, Freiburg, Germany

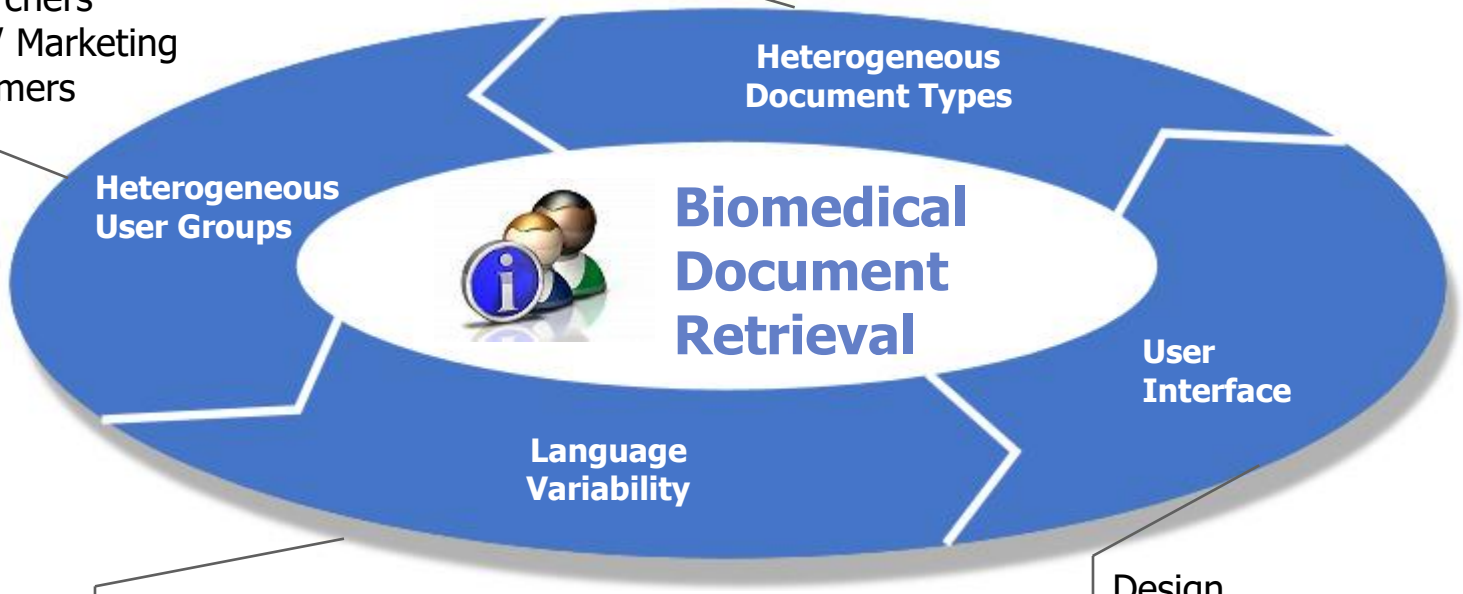
- Cross-language document retrieval in life sciences and health care
- The technique of morphosemantic document indexing
- Evaluation of morphosemantic indexing
- Cross-language document retrieval in practice

- Cross-language document retrieval in life sciences and health care
- The technique of morphosemantic document indexing
- Evaluation of morphosemantic indexing
- Cross-language document retrieval in practice

Biomedical Documents

Electronic Patient Records
Textbook Information
Product Information
Experimental Reports
Scientific Publications
Websites

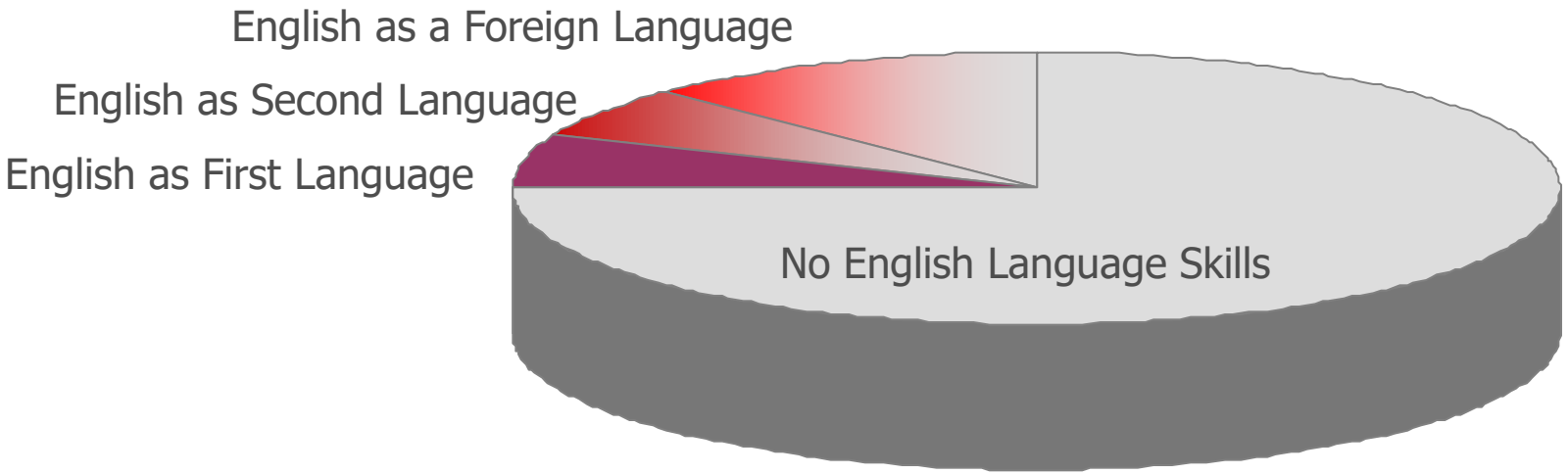
Health Professionals
Researchers
Sales / Marketing
Consumers



Laypersons' language
Experts' language
Subdomain-related jargon
Language in the lab / in scientific publication
Language of the health record
Writer's / Reader's native / 2nd / foreign language

Design
Accessibility
Bridging the gap between user groups and information sources

The Status of English Language



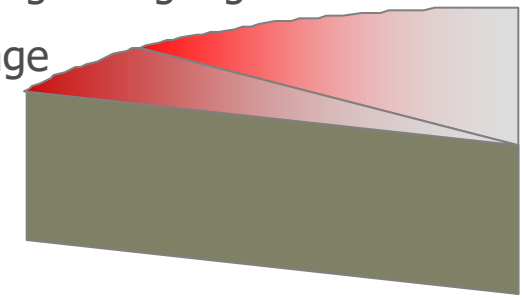
- < 70 % of the world's scientists *read* in English
- 80 % of the world's electronically stored information is in English
- 90 % English articles in Medline (2000)

Sources: The British Council, 2005
Fung ICH: Open access for the non-English-speaking world: overcoming the language barrier. Emerging Themes in Epidemiology, 2008

Non-native speakers

English as a Foreign Language

English as Second Language



- Broad range of command of English
- Reading skills > writing skills
- Reduced active vocabulary

→ Difficulty in formulating precise queries

Cross-language IR

- Cross language information retrieval (CLIR) deals with retrieving information written in a language different from the language of the user's query
- Benefit for multilingual users
  - Avoid multiple queries
  - Formulate a query in their preferred language

- Monolingual users take advantage
  - if their passive knowledge is sufficient to understand documents in a foreign language
  - If (automatic) translation can be performed
  - If image captions are used to search for images

CLIR Beyond Native -> English

- Mixed document collections (in different languages)
- Countries with more than one official language (e.g. Switzerland, Canada, Belgium, Spain…)
- Document handsearching, e.g. Freiburg Cochrane Collaboration project (since 1995)
  - Identification of 21,620 controlled clinical trials
  - 83% not listed in MEDLINE as „controlled trial"
  - 30% not indexed in MEDLINE
  - 30% not in English language

averbis
medical language technology

Example

UNIVERSITÄTS KLINIKUM FREIBURG

Korrelation von Hypertonie und Läsion der Weißen Substanz…

Example

Korrelation von Hypertonie und Läsion der Weißen Substanz…

"Correlation of high blood pressure and lesion of the white substance"

Interaction between hypertension, apoE, and cerebral white matter lesions.

de Leeuw FE, Richard F, de Groot JC, van Duijn CM, Hofman A, Van Gijn J, Breteler MM.

Department of Epidemiology and Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands.

BACKGROUND AND PURPOSE: Cerebral white matter lesions (WMLs) are frequently found on magnetic resonance imaging scans in both cognitively intact and demented elderly persons. Vascular risk factors, especially hypertension, are related to their presence. However, not every person with vascular risk factors has WMLs, which suggests interaction with other determinants, eg, genetic factors. The epsilon4 allele

averbis
medical language technology

Example

UNIVERSITÄTS KLINIKUM FREIBURG

Korrelation von Hypertonie und Läsion der Weißen Substanz…

"Correlation of high blood pressure and lesion of the white substance"

Interaction between hypertension, apoE, and cerebral white matter lesions.

de Leeuw FE, Richard F, de Groot JC, van Duijn CM, Hofman A, Van Gijn J, Breteler MM.

Department of Epidemiology and Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands.

BACKGROUND AND PURPOSE: Cerebral white matter lesions (WMLs) are frequently found on magnetic resonance imaging scans in both cognitively intact and demented elderly persons. Vascular risk factors, especially hypertension, are related to their presence. However, not every person with vascular risk factors has WMLs, which suggests interaction with other determinants, eg, genetic factors. The epsilon4 allele

Example

Korrelation von Hypertonie und Läsion der Weißen Substanz…

"Correlation of high blood pressure and lesion of the white substance"

Interaction betw en hypertension, apoE, and cerebral white ma r lesions.

de Leeuw FE, Richard F, Groot JC, van ijn CM, Hofman A, Van Gijn J, Bret er MM.

Department of Epidemiology and st istics, Erasmus Medical Center, Rotterdam, The Net lands.

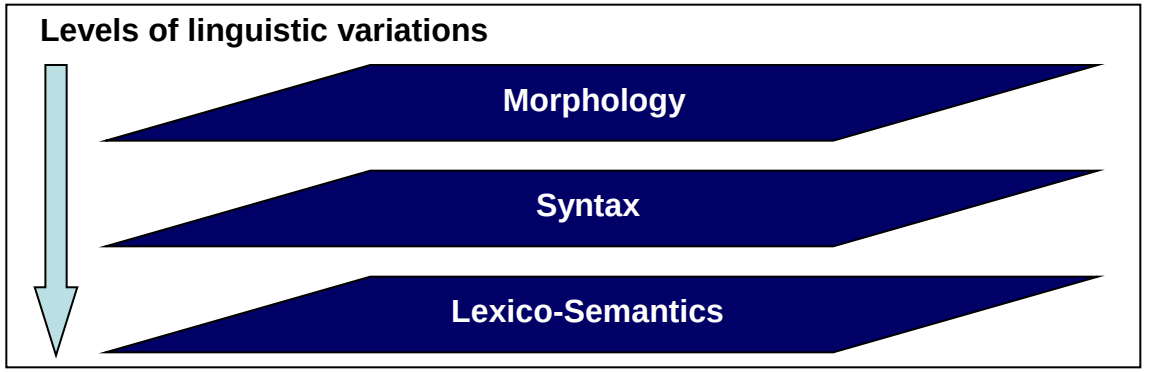
BACKGROUND AND PURPO E: Cerebr white matter lesions (WMLs) are frequent found on ma etic resonance imaging scans in both cogn vely intact and de nted elderly persons. Vascular risk fa ors, especially hyperte ion, are related to their presen . However, not every perso ith vascular risk factor as WMLs, which suggests intera on with other determ ants, eg, genetic factors. The epsilon4 llele

- Cross-language document retrieval in life sciences and health care
- The technique of morphosemantic document indexing
- Evaluation of morphosemantic indexing
- Cross-language document retrieval in practice

Hypotheses

- The true, significant elements of language are . . . either words, significant parts of words, or word groupings. [Sapir 1921]

- Linguistic variations make (medical) Information Retrieval difficult



Morphosemantic Indexing

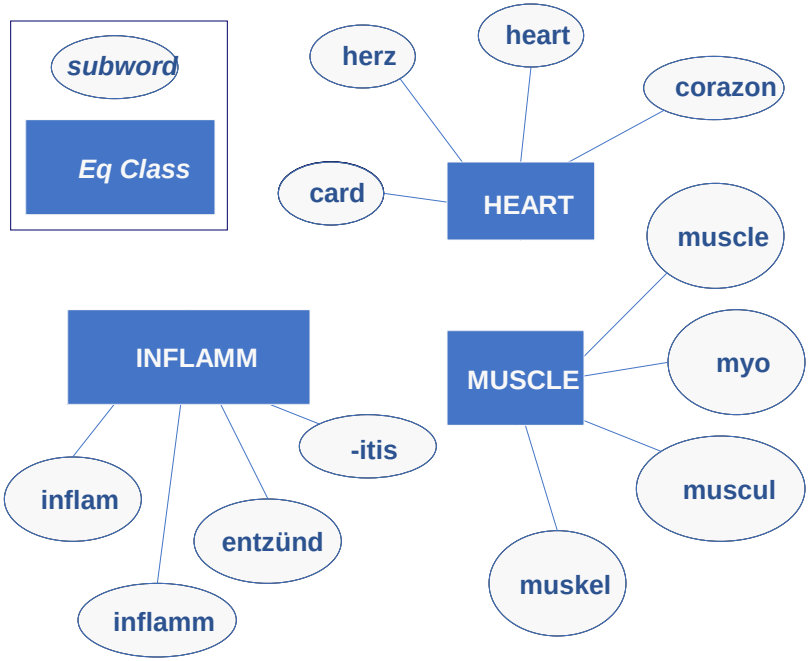
- Subword-based, multilingual semantic indexing for document retrieval

- Subwords are atomic, conceptual or linguistic units:
  - Stems: stomach, gastr, diaphys
  - Prefixes: anti-, bi-, hyper-
  - Suffixes: -ary, -ion, -itis
  - Infixes: -o-, -s-

- Equivalence classes contain synonymous subwords and their translations:
  - **#derma = { derm, cutis, skin, haut, kutis, pele, cutis, piel, ... }**
  - **#inflamm = { inflamm, -itic, -itis, -phlog, entzuend, -itis, -itisch, inflam, flog, inflam, flog, ... }**

Subword Thesaurus Structure

- Thesaurus:
  ~21.000 equivalence classes (MIDs)
- Lexicon entries:
  - English: ~23.000
  - German: ~24.000
  - Portuguese: ~15.000
  - Spanish : ~11.000
  - French: ~ 8.000
  - Swedish: ~10.000
  - Italian: ~ 4.000



Segmentation:

Myo | **kard** | **itis**

Herz | **muskel** | **entzünd** | ung

Inflamm | ation of the **heart muscle**

Indexation:

#muscle #heart #inflamm

#heart #muscle #inflamm

#inflamm #heart #muscle

Indexing Pipeline

Original Document
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.

Indexing Pipeline

Original Document	Orthographic Normalization
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.	high tsh values suggest the diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.	erhoehte tsh-werte erlauben die diagnose einer primaeren hypothyreose, ein supprimierter tsh-spiegel spricht dagegen fuer eine schilddruesenueberfunktion.
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.	a presenca de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.

Indexing Pipeline

Original Document	Orthographic Normalization	Morphological Segmentation
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.	high tsh values suggest the diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.	high tsh value s suggest the diagnos is of primar y hypo thyroid ism while a suppress ed tsh level suggest s hyper thyroid ism.
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.	erhoehte tsh-werte erlauben die diagnose einer primaeren hypothyreose, ein supprimierter tsh-spiegel spricht dagegen fuer eine schilddruesenueberfunktion.	er hoeh te tsh - wert e erlaub en die diagnos e einer primaer en hypo thyre ose, ein supprim iert er tsh - spiegel spricht dagegen fuer eine schilddrues en ueber funktion.
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.	a presenca de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.	a presenc a de valor es elevad os de tsh suger e o diagnost ico de hipo tireoid ismo primari o, enquanto niveis suprimid os de tsh suger em hiper tireoid ismo.

Indexing Pipeline

Original Document	Orthographic Normalization	Morphological Segmentation	Semantic Normalization
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.	high tsh values suggest the diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.	high tsh value s suggest the diagnos is of primar y hypo thyroid ism while a suppress ed tsh level suggest s hyper thyroid ism.	**#up# tsh #value#** #suggest# **#diagnost# #primar# #small# #thyre# #suppress# tsh #nivell#** #suggest# **#up# #thyre#** .
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.	erhoehte tsh-werte erlauben die diagnose einer primaeren hypothyreose, ein supprimierter tsh-spiegel spricht dagegen fuer eine schilddruesenueberfunktion.	er hoeh te tsh - wert e erlaub en die diagnos e einer primaer en hypo thyre ose, ein supprim iert er tsh - spiegel spricht dagegen fuer eine schilddrues en ueber funktion.	**#up# tsh - #value#** #permit# **#diagnost# #primar# #small# #thyre#** , **#suppress# tsh** - {#mirrorr# **#nivell#**} #speak# **#thyre# #up#** #function# .
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.	a presenca de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.	a presenc a de valor es elevad os de tsh suger e o diagnost ico de hipo tireoid ismo primari o, enquanto niveis suprimid os de tsh suger em hiper tireoid ismo.	#actual# **#value# #up# tsh** #suggest# **#diagnost# #small# #thyre# #primar#** , **#nivell# #suppress# tsh** #suggest# **#up# #thyre#** .

Subword-based document transformation

Interaction between hypertension, apoE, and cerebral white matter lesions.

de Leeuw FE, Richard F, de Groot JC, van Duijn CM, Hofman A, Van Gijn J, Breteler MM.

Department of Epidemiology and Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands.

BACKGROUND AND PURPOSE: Cerebral white matter lesions (WMLs) are frequently found on magnetic resonance imaging scans in both cognitively intact and demented elderly persons. Vascular risk factors, especially hypertension, are related to their presence. However, not every person with vascular risk factors has WMLs, which suggests interaction with other determinants, eg, genetic factors. The epsilon4 allele

Morphosemantic indexer

#interact #hyper #tens , apoe , #cerebr #whit #matter #lesion .

de leeuw fe , richard f , de groot jc , van duijn cm , hofman a , van gijn j , breteler mm .

#department #epidem #logic #bio #statist , erasmus #medic #centr , rotterdam , #dutch .

#back #ground #purpos : #cerebr #whit #matter #lesion (wmls) #frequent #find #magnet #resonanc #imag #scan #both #cognit #intact #dement #gero #human . #vascul #risk #factor , #special #hyper #tens , #relat #presenc . #not #total #human #vascul #risk #factor wmls ,# suggest #interact #other #determin, eg ,

Subword-Based Search

Korrelation von Hypertonie und Läsion der Weißen Substanz…

#correl #hyper #tens #lesion #whit #matter

#interact #hyper #tens , apoe , #cerebr #whit #matter #lesion .

de leeuw fe , richard f , de groot jc , van duijn cm , hofman a , van gijn j , breteler mm .

#department #epidem #logic #bio #statist , erasmus #medic #centr , rotterdam , #dutch .

ck #ground #purpos : #cerebr #whit #matter #lesion (wmls) uent #find #magnet #resonanc #imag #scan #both #cognit ct #dement #gero #human . #vascul #risk #factor , #special yper #tens , #relat #presenc . #not #total #human #vascul #risk #factor wmls ,# suggest #interact #other #determin, eg ,

Subword-based query transformation

Korrelation von Hypertonie und Läsion der Weißen Substanz…

#correl #hyper #tens #lesion #whit #matter

#interact #hyper #tens , apoe , #cerebr #whit #matter #lesion .

de leeuw fe , richard f , de groot jc , van duijn cm , hofman a , van gijn j , breteler mm .

#department #epidem #logic #bio #statist , erasmus #medic #centr , rotterdam , #dutch .

ck #ground #purpos : #cerebr **#whit #matter #lesion** (wmls quent #find #magnet #resonanc #imag #scan #both #cognit t #dement #gero #human . #vascul #risk #factor , #special per **#tens** , #relat #presenc . #not #total #human #vascul risk #factor wmls ,# suggest #interact #other #determin, eg ,

- Cross-language document retrieval in life sciences and health care
- The technique of morphosemantic document indexing
- Evaluation of morphosemantic indexing
- Cross-language document retrieval in practice

Evaluation

- Gold standards: OHSUMED, ImageCLEFMed
  - OHSUMED-Corpus (Hersh et al., 1994)
    - Subset of MEDLINE
    - ~233,000 English documents
    - 106 English user queries
  - ImageCLEFMed Corpus (Clough et al., 2005)
    - Multilingual Image Retrieval Task 2006
    - ~41.000 Medical Images and captions
    - 30 queries
- Query-document pairs had been manually judged for relevance
- Non-English queries were obtained by translation to German, Portuguese, Spanish and Swedish by domain experts
- Search Engine: Lucene
  - http://lucene.apache.org/

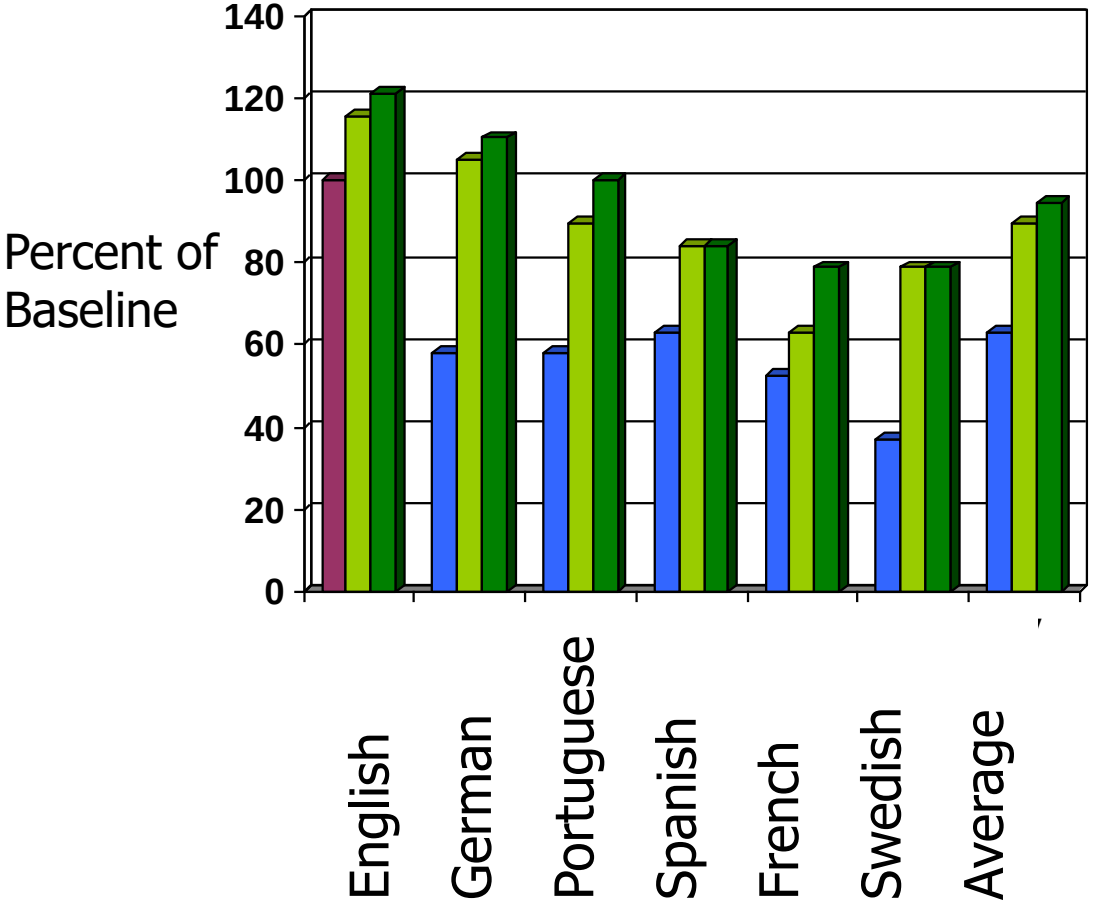
averbis
medical language technology

Evaluation

UNIVERSITÄTS KLINIKUM FREIBURG

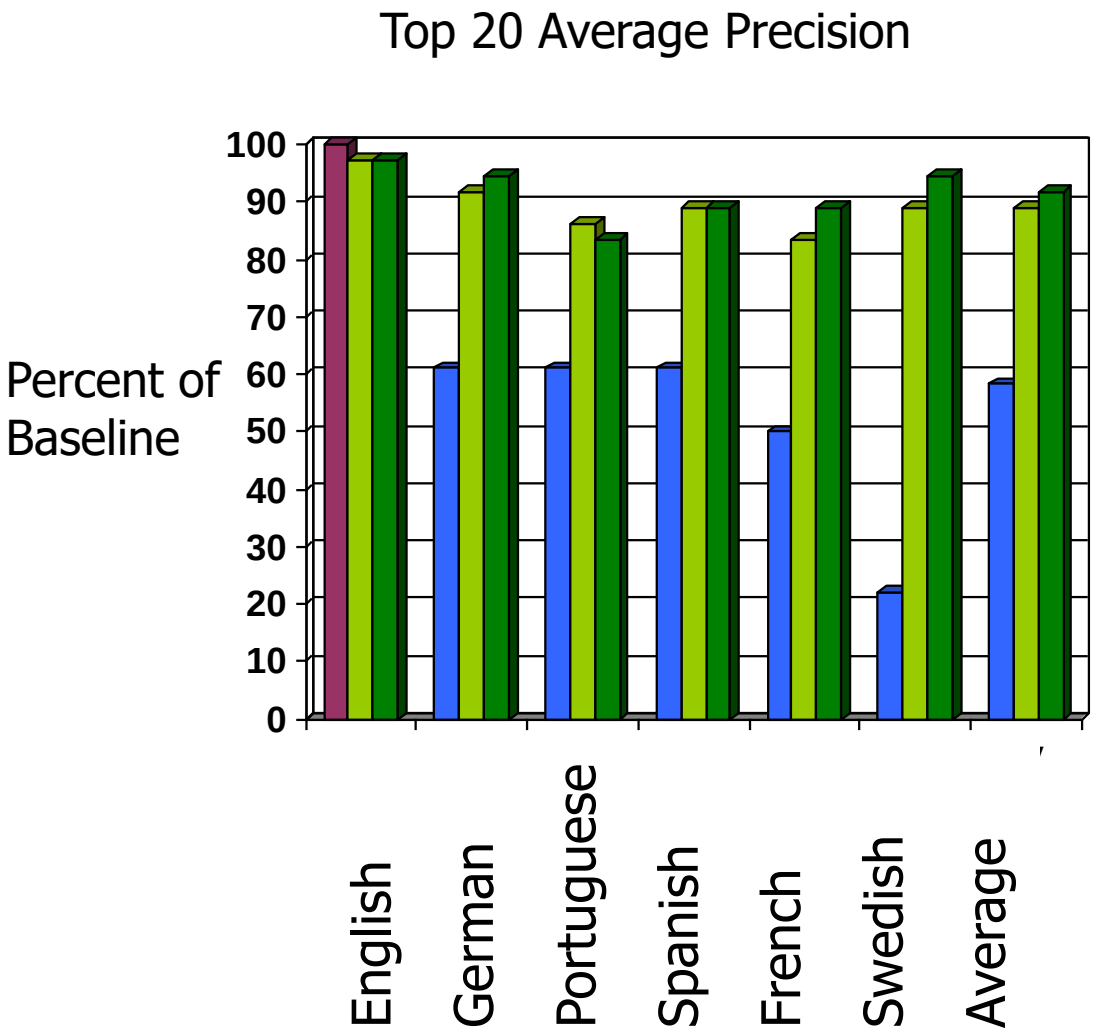
- Baseline: monolingual text retrieval
  - (stemmed) English user queries
  - (stemmed) English texts
- Query translation (QTR)
  - Google translator
  - Multilingual dictionary compiled from UMLS
- Morphosemantic Indexing (MSI)
  - Interlingual representation of both user queries and documents
  - MSI-D incorporating disambiguation module

OHSUMED Benchmark



- **Baseline**: monolingual
  - Stemmed English queries
  - Stemmed English texts
- **Query translation**
  - Google translator
  - Multilingual dictionary compiled from UMLS
- **Morphosemantic Indexing**
  - Interlingual representation of user queries and documents
- **Morphosemantic Indexing**
  - incorporating disambiguation module

ImageCLEFMed Benchmark



- **Baseline**: monolingual
  - Stemmed English queries
  - Stemmed English texts
- **Query translation**
  - Google translator
  - Multilingual dictionary compiled from UMLS
- **Morphosemantic Indexing**
  - Interlingual representation of user queries and documents
- **Morphosemantic Indexing**
  - incorporating disambiguation module

Summary

- Cross-Language Document Retrieval
  - Based on morphological and semantic normalization of both user queries and documents
  - Matching of search/document terms on a language-independent, interlingual layer
- Language-independent indexing
  - reaches more than **92%** of an English-English baseline on heterogeneous document collections, on average
  - outperforms query translation significantly
  - is independent from particular search engine architectures
- Incorporates six languages:
  - German, English, Portuguese, Spanish, French, Swedish
- In use in commercial systems

- Cross-language document retrieval in life sciences and health care
- The technique of morphosemantic document indexing
- Evaluation of morphosemantic indexing
- Cross-language document retrieval in practice

averbis
medical language technology

Zentralbibliothek für Medizin

UNIVERSITÄTS KLINIKUM FREIBURG

- Largest European medical library
- ~20 M Database entries
- 60,000 Queries / Month
- Subword-based cross-language retrieval

MEDPILOT.DE
Eine Seite. Alles Wissen.

Suche | Suchergebnisse | Merkliste [0] | Dokumentbestellung

ösophagusvarizenblutung [Suchen]

Suchergebnisse (neu)

Trefferzahlen der einzelnen Datenbanken

Hogrefe Verlag	2
Karger Verlag	0
Kluwer Verlag	0
Krause und Pachernegg Publikations-Datenbank	0
MEDLINE	0
Springer Verlag	3
Thieme Verlag	4
Katalog ZB MED Medizin	29
Katalog ZB MED Ernährung / Umwelt / Agrar	0
CCMED	15
Treffer insgesamt	**53**

Mehr Datenbanken und Optionen

Literaturagenten einrichten

Datenbank: **Hogrefe Verlag** Treffer: 2

Treffer: 1-2 Aktuell: 1-2 zeigen

Complications of Liver Cirrhosis: Portal Hypertension, Gastroesophageal Varico
Komplikationen der Leberzirrhose: Portale Hypertension, gastroösophageale V
Schuster MJ {a}
PRAXIS
2003 / 92 (35) 1427-1434
Patients with cirrhosis of the liver are at high risk of a large variety of complications. Espe portal hypertension, followed by gastroesophageal varicosis and ascites are potentially life the treatme...

☐ Merkliste

Terlipressin bei akuter Ösophagusvarizenblutung
Ioannou G; Doust J; Rockey DC
PRAXIS
2002 / 91 (20) 901-901

☐ Merkliste

Treffer: 1-2 Aktuell: 1-2 zeigen

sichere Verbindung | lokale Sichten | Impressum/Datenschutz | Kontakt | MEDPILOT.DE auf der eigenen

averbis
medical language technology

UNIVERSITÄTS KLINIKUM FREIBURG

ösophagusvarizenblutung | Go | Sort By: RELEVANCE

Deutsche Version

Found 405 Results in 642 msecs.

Disease
- Varicose Veins
- Esophageal and Ga...
- Hemorrhage
- Gastrointestinal ...
- Liver Cirrhosis

Investigations
- Esophagoscopy
- Ultrasonography, ...
- Tomography, X-Ray...
- Endoscopy, Gastro...
- Phlebography

Therapy
- Vascular Surgical...
- Sclerotherapy
- Blood Transfusion
- Drug Therapy, Com...
- Ligation

Anatomy
- Saphenous Vein
- Leg
- Lung
- Portal Vein
- Esophagus

Chemicals
- Sclerosing Soluti...
- Adrenergic beta-A...
- Anticoagulants
- Polyethylene Glyc...
- Warfarin

Journal
- Gastrointestinal ...
- European journal ...

Savel'ev, V S ; Prokubovskiĭ, V I ; Ovchininskiĭ, M N ; Cherkasov, V A: view in PubMed

[Intravascular occlusion in hemorrhages from varicose gastric and esophageal veins]

... system of the superior caval vein through the **varicose** dilated **esophageal** veins is known to be... of the most effective methods to stop profuse **bleedings** due_to a disturbed integrity of the **varicose**...

Keywords:Catheterization; Embolization, Therapeutic; **Esophageal** and Gastric Varices; Gastrointestinal **Hemorrhage**; Humans; Portal Vein; Stomach; **Varicose** Veins

Vestnik khirurgii imeni I. I. Grekova, 1983, 130(5), pp. 29-33, ISSN: 0042-4625

Sauerbruch, T ; Weinzierl, M ; Dietrich, H P ; Antes, G ; Eisenburg, J ; Paumgartner, G: view in PubMed

Sclerotherapy of a bleeding duodenal varix.

A case of successful treatment of a **bleeding** duodenal **varix** in a patient with portal hypertension... 42-year-old man had a history of recurrent gastrointestinal **hemorrhage** over 14 years. In 1966 he underwent... Splenectomy was performed because of hypersplenism. In 1980 **bleeding** from **esophageal** varices occurred and was treated... weeks after sclerotherapy massive **bleeding** from a duodenal **varix** occurred. Sclerotherapy of the duodenal **varix** via...

Keywords:Adult; Duodenoscopy; Duodenum; Gastrointestinal **Hemorrhage**; Humans; Male; Polyethylene Glycols; Sclerosing Solutions; **Varicose** Veins

Endoscopy, 1982, 14(5), pp. 187-9, ISSN: 0013-726X

Onopriev, V I ; Durleshter, V M ; Usova, O A ; Kliuchnikov, O Iu: view in PubMed

[Surgical treatment of bleedings from varicose veins of the esophagus and stomach]

Experience in surgical treatment of **bleedings** from **varicose** veins of the **esophagus** and stomach is analyzed....

Keywords:Adult; **Esophageal** and Gastric Varices; **Esophagus**; Female; Gastrointestinal **Hemorrhage**; Gastroplasty; Humans; Male; Vagotomy, Proximal Gastric

Khirurgiia, 2005, (1), pp. 38-42, ISSN: 0023-1207

Solomatin, A D: view in PubMed

[Hemostasis in hemorrhage from varicose veins of the esophagus and stomach]

Contact

Prof. Dr. med. Stefan Schulz

Institute of Medical Biometry and Medical Informatics
University Medical Center
Freiburg, Germany

stschulz@uni-freiburg.de http://www.imbi.uni-freiburg.de

Dr. med. Philipp Daumke

Averbis GmbH
Freiburg, Germany

daumke@averbis.de http://www.averbis.net