HISA ltd. Biography proforma

MEDINFO 2007

413 Lygon Street, Brunswick East 3057 Australia

Presenter Name: Stefan Schulz
Country: 1. Germany, 2. Brazil
Qualification(s):

MD (Doctor in Theoretical Medicine)
Vocational Training in Medical Informatics
Postdoctoral Habilitation degree in Medical Informatics

Position:

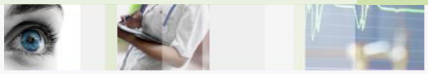
Associate Professor

Department/ Organisation :

1. Medical Informatics, Freiburg University Medical Center, Freiburg, Germany
2. Master Program of Health Technology, Catholic University of Paraná, Curitiba, Brazil

Major Achievement(s):

Research in the fields of Medical Terminology, Biomedical Ontologies, Medical Language Processing, Text Mining, and Document Retrieval.

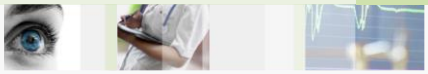


Large-Scale Evaluation of a Medical Cross-Language Information Retrieval System

Kornél Markó[1,2], Philipp Daumke[1,2], Stefan Schulz[2],
Rüdiger Klar[2], Udo Hahn[3]

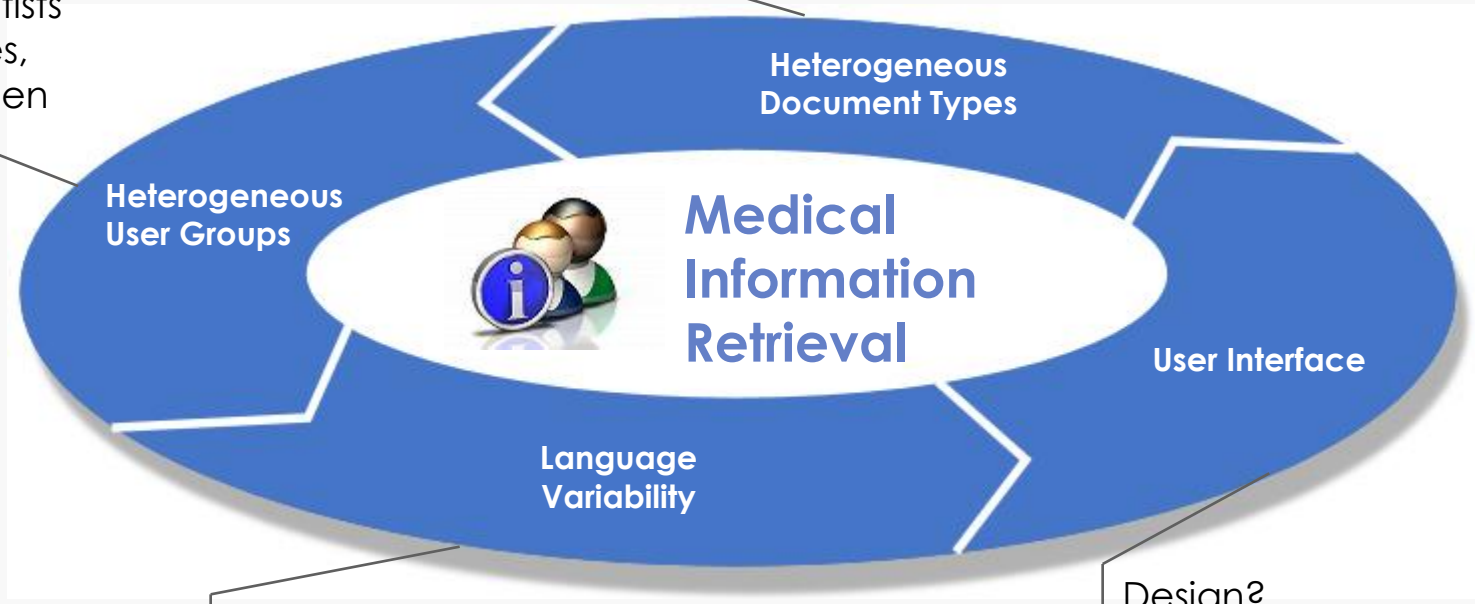
[1]Averbis GmbH, Freiburg, Germany
[2]Medical Informatics Department, University Medical Center Freiburg, Germany
[3]Jena University Language and Information Engineering Lab (JULIE), Germany

Medical Information



Discharge Summaries
Pathology Reports,
Scientific Publications, etc.

Physicians,
Scientists
Nurses,
Laymen



Experts vs. Laymen
Linguistic Morphology
Cross-Lingual

Design?
Accessibility?
Bridging the gap
between user groups and
information sources?

User Interface

Risk factor high blood pressure

Search for risk factor high blood pressure

Risk factor**s** hypertension

Search for risk factors hypertension

Risk factor hypertension

Search for risk factor hypertension

User Interface

averbis
medical language technology

Risk factor high blood pressure

Risk factors hypertension

Risk factor hypertension

Items 1 - 20 of 31298

Schuhlen H, Abts M, Kast

[Intensive Blood ssure Reduction in Patients with Increased Cardiovascular Risk with High-Dose Combi
Hydrochlorothiazide. Results of the MACHT II Observational Study.]
Herz. 2007 Jul;32(5):419-25. German.

Items 1 - 20 of 18656

Schuhlen H, Abts M, Ka

[Intensive Blood ssure Reduction in Patients with Increased Cardiovascular Risk with High-Dose C
Hydrochlorothiazide. Results of the MACHT II Observational Study.]
Herz. 2007 Jul;32(5):419-25. German.

Items 1 - 20 of 28255

: Schuhlen H, Abts M, Ka

[Intensive Blood ssure Reduction in Patients with Increased Cardiovascular Risk with High-Dose
Hydrochlorothiazide. Results of the MACHT II Observational Study.]
Herz. 2007 Jul;32(5):419-25. German.

UNIVERSITÄTS FREIBURG KLINIKUM

Multilingualism

averbis
medical language technology

Korrelation von Hypertonie und Läsion der Weißen Substanz…

UNIVERSITÄTS FREIBURG KLINIKUM

Multilingualism

averbis
medical language technology

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

Multilingualism

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

Multilingualism

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 deter ants, eg, genetic factors. The epsilon4 llele

Linguistic Morphology

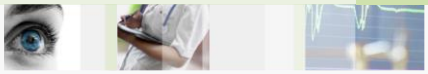
Linguistic phenomena adversely influence medical text retrieval !

- **Inflection**: leukocyte vs. leukocytes, appendix vs. appendices
- **Derivation**: leukocyte, vs. leukocytic
- **Composition**: leuk | em | ia, para | sympath | ectomy
  Magen | schleim | haut | entzünd | ung
- **Acronyms**: AIDS, SARS, OECD
- **Orthographic Variants**:
  - Kolonkarzinom, Coloncarcinom,
  - Esophagus, Oesophagus,
- **Synonyms**:
  - High blood pressure – Hypertension,
  - Prophylaxis – Prevention

UNIVERSITÄTS KLINIKUM FREIBURG

MorphoSaurus

averbis
medical language technology



Subword-based, multilingual semantic indexer for document retrieval

Subwords are atomic, conceptual or linguistic units:

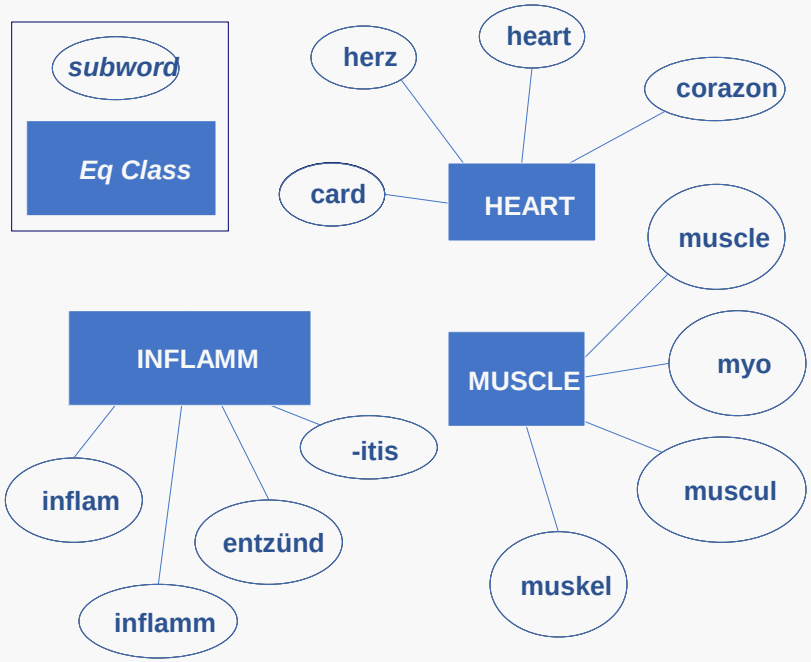
- Stems: *stomach, gastr, diaphys*
- Prefixes: *anti-, bi-, hyper-*
- Suffixes: *-ary, -ion, -itis*
- Infixes: *-o-, -s-*

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, … }**

Morphosaurus 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



Segmentation:

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

Herz | **muskel** | **entzünd** | ung

Inflamm | ation of the **heart muscle**

Indexation:

#muscle #heart #inflamm

#heart #muscle #inflamm

#inflamm #heart #muscle

UNIVERSITÄTS FREIBURG KLINIKUM

Morpho-Semantic Normalization

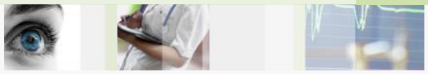
averbis
medical language technology

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 ueber funkti	**#up# tsh #value#** #suggest# **#diagnost# #primar# #small# #thyre#** , **#suppress#** **tsh** - {#mirror# **#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 pr es sugerem hiper 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#** .

UNIVERSITÄTS KLINIKUM FREIBURG

Disambiguation

averbis
medical language technology



- Maximum likelihood estimator
- Co-occurence information from large heterogeneous courpora
- #patient should be preferred over #patience, since „Patient“ is unambiguous in German and also co-occurs with #heart

The American Heart Association recommends aspirin use for any patient who had a heart attack.
In Deutschland leben 120.000 erwachsene Patienten mit angeborenen Herzfehlern.

#america **#heart** #associat #advice #aspirin #utilis {**#patient**, #patience} #heart #attack.
#german #life 120.000 #adult **#patient** #heredit **#heart** #failure.

UNIVERSITÄTS KLINIKUM FREIBURG

Subword-based Search

averbis
medical language technology

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

MorphoSaurus

#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 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 …quent #find #magnet #resonanc #imag #scan #both #cognit …ct #dement #gero #human . #vascul #risk #factor , #special …**per #tens** , #relat #presenc . #not #total #human #vascul …risk #factor wmls ,# suggest #interact #other #determin, eg ,

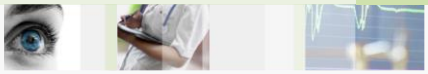
UNIVERSITÄTS KLINIKUM FREIBURG

Evaluation

averbis
medical language technology

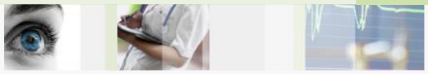
- Gold standards: OHSUMED, ImageCLEFMed
- OHSUMED-Corpus (Hersh et al., 1994)
  - Subset of MEDLINE
  - ~233,000 English documents
  - 106 English user queries
- ImageCLEFMed Corpus (Clough et al., 2005)
  - Multilingual Image Retrieval Task 2006
  - ~41.000 Medical Images and captions
  - 30 queries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

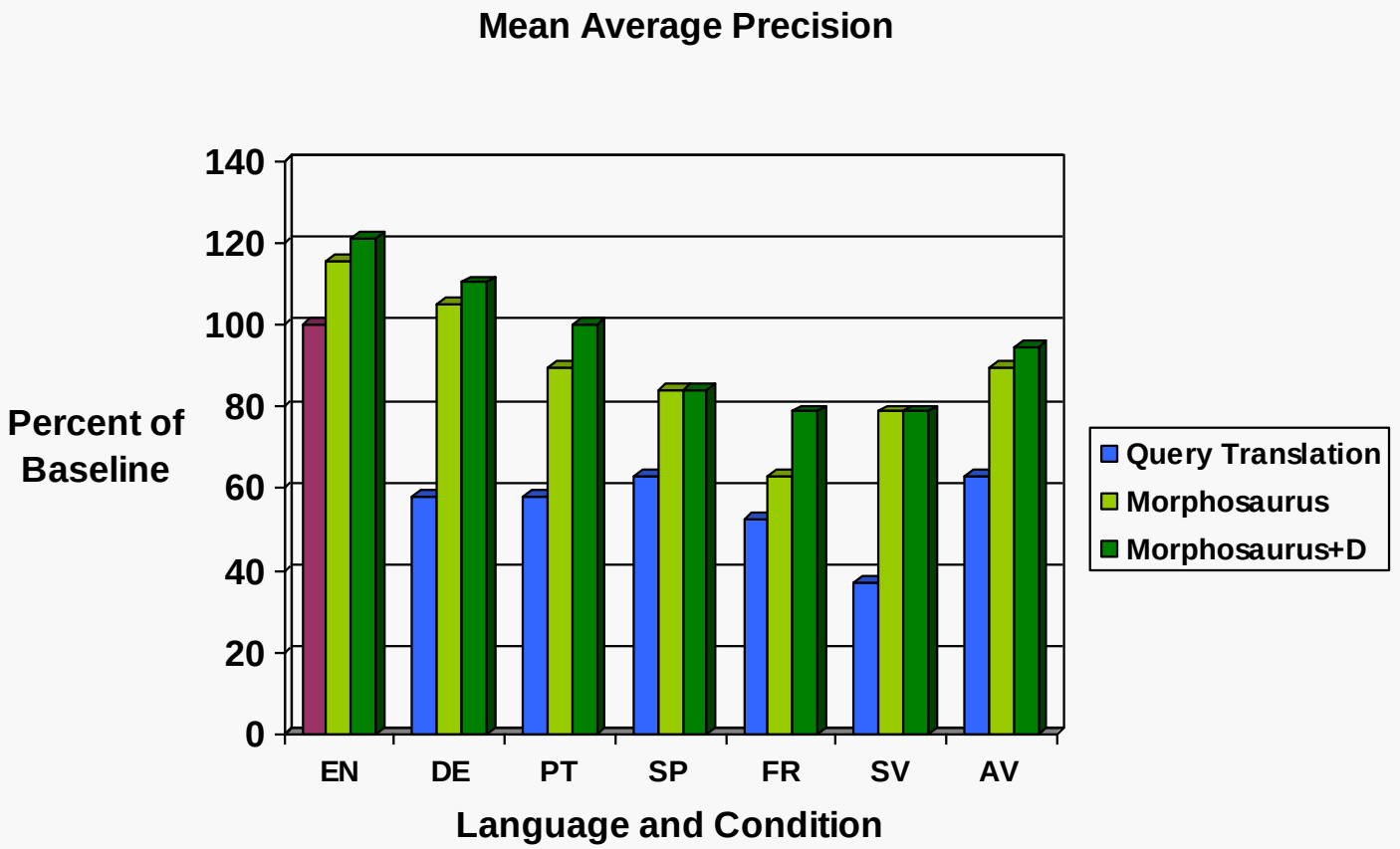
Evaluation



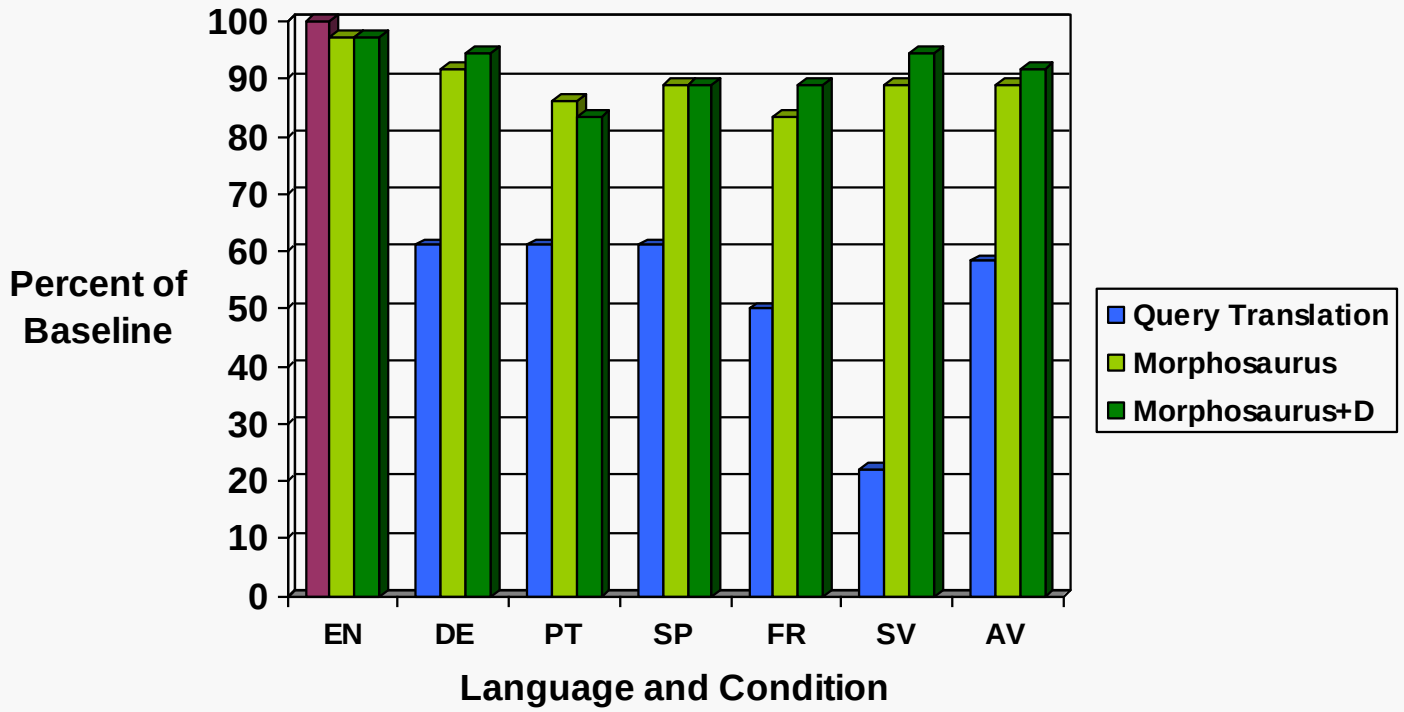
- Baseline: monolingual text retrieval
  - (stemmed) English user queries
  - (stemmed) English texts
- Query translation (QTR)
  - Google translator
  - Multilingual dictionary compiled from UMLS
- Morphosaurus Indexing (MSI)
  - Interlingual representation of both user queries and documents
  - MSI-D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

Results: Ohsumed





Results: ImageCLEFMed



UNIVERSITÄTS FREIBURG KLINIKUM

Conclusions

averbis
medical language technology

- 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, in average
  - outperforms query translation significantly
  - is independent from particular search engine architectures
- Morphosaurus incorporates six languages:
  - German, English, Portuguese, Spanish, French, Swedish
- In use in commercial systems

Contact

Stefan Schulz
Medical Informatics Department
University Hospital Center Freiburg, Germany
stschulz@uni-freiburg.de

www.imbi.uni-freiburg.de

Kornél Markó
Averbis GmbH, Freiburg, Germany
marko@averbis.de

www.averbis.net