

averbis
medical language technology

German – Brazilian Cooperation (BMBF-CNPq)



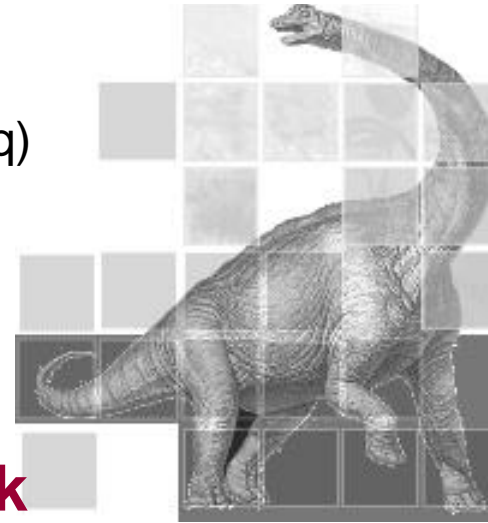
PUCPR



**UNIVERSITÄTS
FREIBURG KLINIKUM**

MorphoSaurus

A Medical Cross-Language Information Retrieval Framework



Edson José Pacheco¹, Píndaro S. Cancian¹, Percy Nohama¹,
Philipp Daumke^{2,3}, Kornel Markó^{2,3}, Stefan Schulz²,

¹Paraná Catholic University, Health Informatics Laboratory, Curitiba, Brazil

²Freiburg University Hospital, Medical Informatics Group, Freiburg, Germany

³AVERBIS GmbH, Freiburg, Germany

Introduction

Methods

Evaluation

Outlook

Summary

Document Retrieval Scenarios

Risk factor high blood pressure

Search for risk factor high blood pressure

Risk factors hypertension

Search for risk factors hypertension

Risk factor hypertension

Search for risk factor hypertension



Document Retrieval Scenarios

Risk factor high
blood pressure

Risk factors
hypertension

Risk factor
hypertension



Items 1 - 20 of 31298

[Schuhlen H, Abts M, Kastner A, et al.](#)

[Intensive Blood Pressure Reduction in Patients with Increased Cardiovascular Risk with High-Dose Combination of Amlodipine and 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, Kastner A, et al.](#)

[Intensive Blood Pressure Reduction in Patients with Increased Cardiovascular Risk with High-Dose Combination of Amlodipine and 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, Kastner A, et al.](#)

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

Multilingualism

Correlação de
hipertensão e
lesões da
substância branca



Multilingualism

Correlação de hipertensão e lesões da substância branca

“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

Correlação de hipertensão e lesões da substância branca

“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

Correlação de hipertensão e lesões da substância branca

“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, Groot JC, van Gijn 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

Linguistic Morphology

Linguistic phenomena adversely influence medical text retrieval

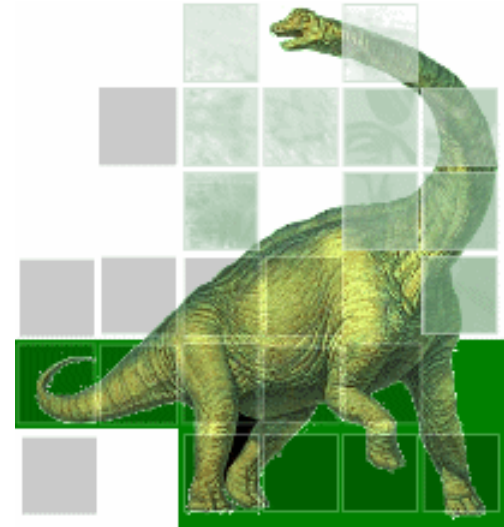
- **Inflection:** leucócit**o** vs. leucócit**os**, append**ix** vs. append**ices**
- **Derivation:** leukocyt**e**, vs. leukocyt**ic**
- **Composition:**
 - leuk|em|ia
 - para|simpat|ectomía
 - Magen|schleim|haut|entzünd|ung
- **Acronyms:** AIDS, SARS, OECD, ECG,
- **Orthographic Variants:**
 - Kolon**k**arzinom, Colon**c**arcin**o**m
 - Esophagus, O**e**sophagus
- **Synonyms:**
 - High blood pressure – Hypertension,
 - Prophylaxis – Prevention

Thesaurus

- Controlled Vocabulary for document indexing and retrieval
- Assigns semantic descriptors (concepts) to (quasi-)synonymous terms
- Contains additional semantic relations (e.g. hyperonym / hyponym)
- Examples: MeSH, UMLS, WordNet
- Multilingual thesaurus: contains translations (cross-language synonymy links)

MorphoSaurus: A medical Thesaurus

- Started in 2000 with two MD theses at Freiburg University
- German-Brazilian Cooperation Funded by BMBF-IB and CNPq since 2001
- Funded by DFG 2003-2006
- Funded by EU SemanticMining NoE 2004-2007, used in EU BOOTStrep 2006-2009
- Spin-Off **averbis** GmbH funded by Baden-Württemberg Federal State
medical language technology
- Industry-strength implementations for a eHealth portal and for the German Central Library of Medicine
- Successful Pilot Application in Freiburg University Hospital (Dermatology)
- Several PhD Theses in Germany and Brazil



MorphoSaurus
(www.morphosaurus.net)

Introduction

Methods

Evaluation

Outlook

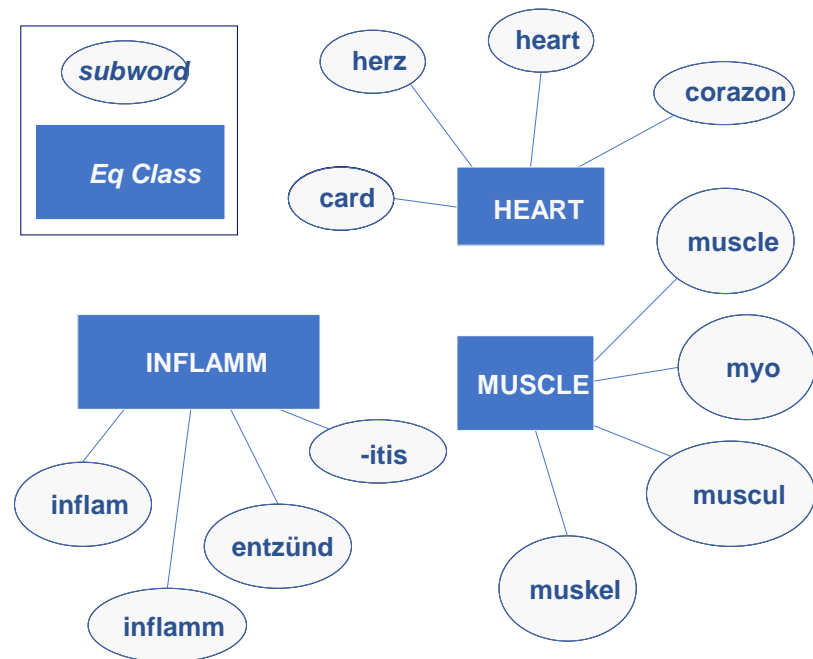
Summary

MorphoSaurus

- Subword-based, multilingual semantic indexer for document retrieval
- Subword = minimum lexical unit for a meaningful term in a domain.
 - Stems: *stomach, gastr, diaphys*
 - Prefixes: *anti-, bi-, hyper-*
 - Suffixes: *-ary, -ion, -itis*
- Equivalence classes contain synonymous subwords and 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 musc

Indexation:

#muscle #heart #inflamm

#heart #muscle #inflamm

#inflamm #heart #muscle

Morpho-Semantic Normalization

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 diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.	#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.	erhoeh te 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 diagnosis einer primaeren hypothyre ose, ein supprim iert er tsh - spiegel spricht dagegen fuer eine schilddruesenueber funktion.	#up# #value# #mit# #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 presenca de valores elevados de tsh sugerem o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.	#actual# #value# #up# tsh #suggest# #diagnost# #small# #thyre# #primar# , #nivell# #suppress# tsh #suggest# #up# #thyre# .

Disambiguation

- Maximum likelihood estimator
- Co-occurrence information from large heterogeneous corpora
- **#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.

Subword-based Search

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

#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 ,



MorphoSaurus

Subword-Based Search

Correlação de hipertensão e lesões da substância branca

#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

Correlação de hipertensão e lesões da substância branca

#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 ,



Introduction

Methods

Evaluation

Outlook

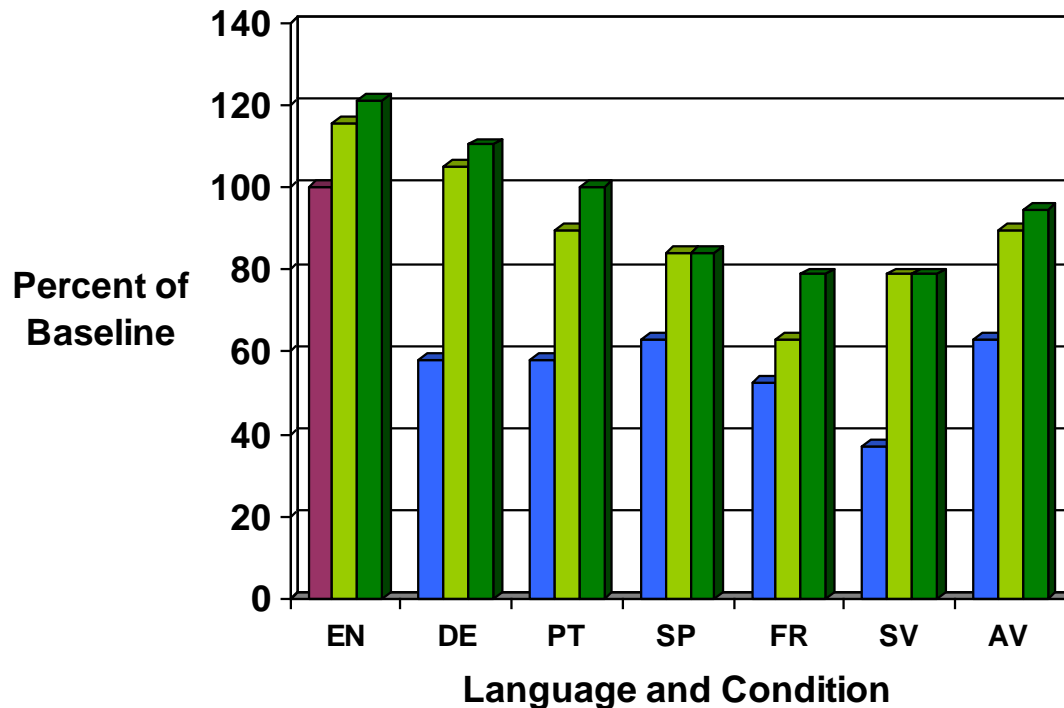
Summary

Evaluation

- Gold standards: OHSUMED
- OHSUMED-Corpus (Hersh et al., 1994)
 - Subset of MEDLINE
 - ~233,000 English documents
 - 106 English user queries
- Query-document pairs have been manually judged for relevance
- Non-English queries obtained by manual translation to German, Portuguese, Spanish and Swedish
- Search Engine: Lucene (<http://lucene.apache.org/>)

Results

Mean Average Precision



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

Introduction

Methods

Evaluation

Outlook

Summary

Further MorphoSaurus Projects

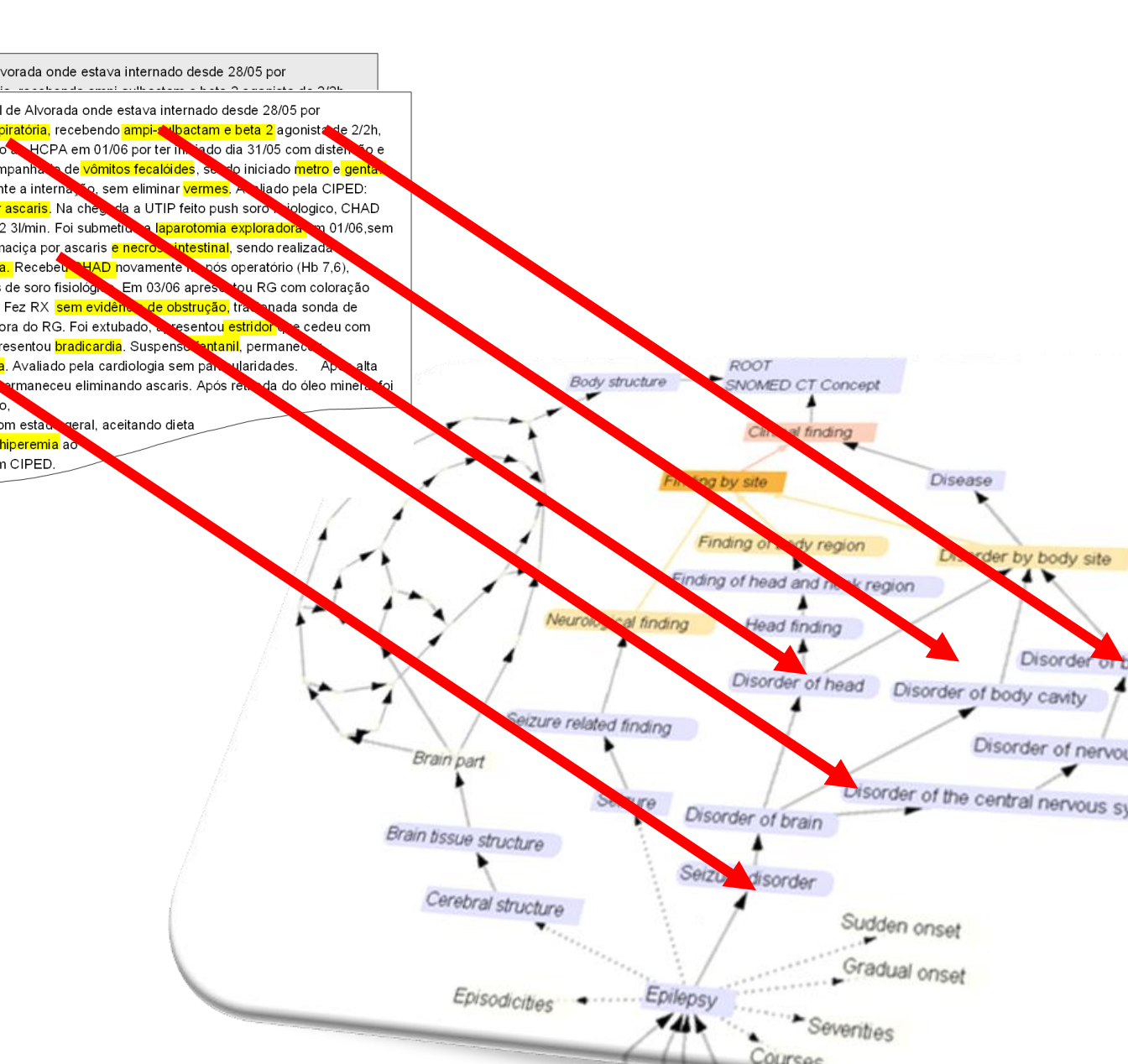
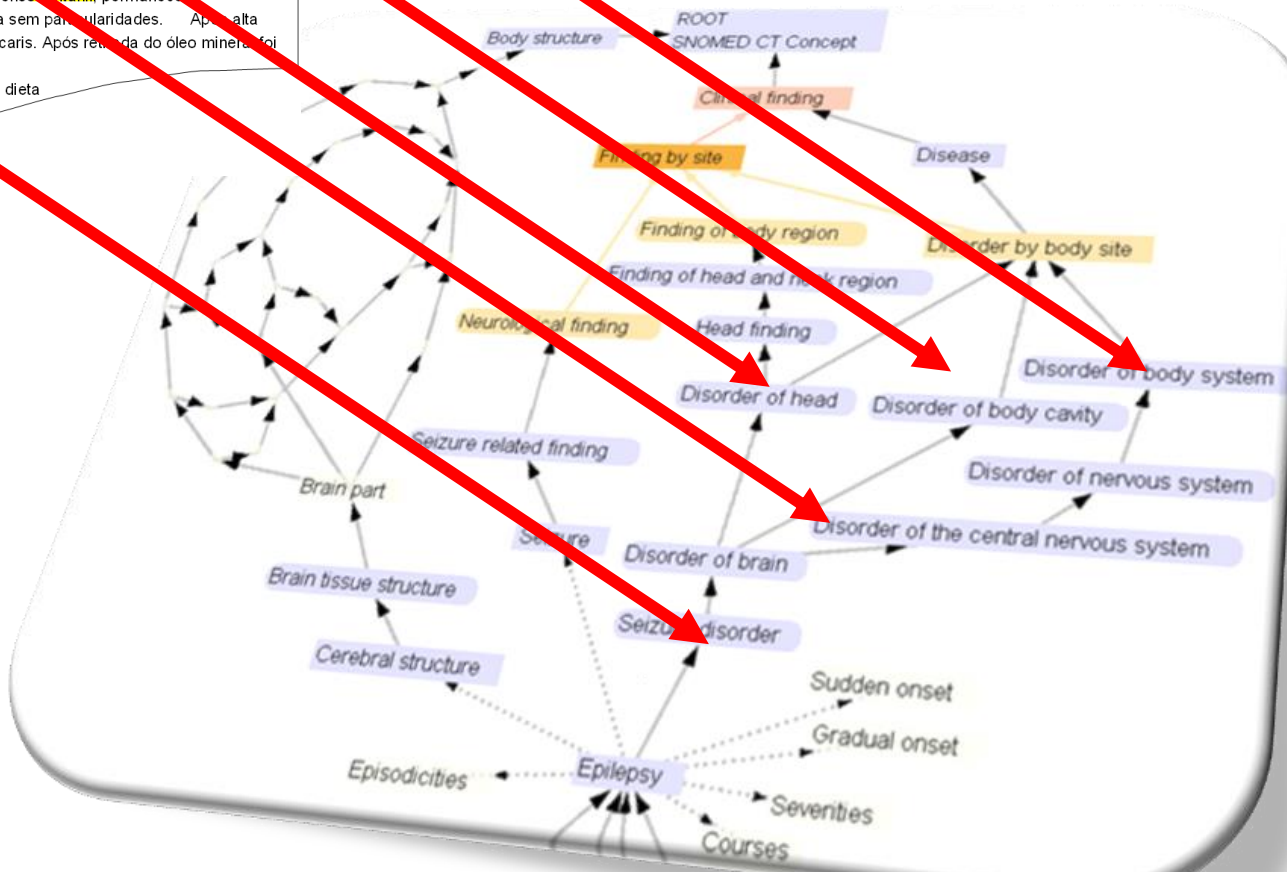
- Retrieval Interface for a lexicon of Molecular Biology (EU BOOTStrep)
- Semantic indexing of German Medical Records for Improving Discharge Diagnoses
- Processing of Hospital Discharge Summaries for automated Ontology mapping (SNOMED CT)

Paciente procedente do Hospital de Alvorada onde estava internado desde 28/05 por broncoespasmo e disfunção respiratória, recebendo ampi-sulbactam e beta 2 agonista de 2/2h, com boa evolução. Encaminhado ao HCPA em 01/06 por ter iniciado dia 31/05 com distensão e dor abdominal importantes, acompanhado de vômitos fecalóides, sendo iniciado metro e genta. Havia recebido albendazol durante a internação, sem eliminar vermes. Avaliado pela CIPED: provável obstrução intestinal por ascaris. Na chegada a UTIP feito push soro fisiológico, CHAD (Hb 9 do dia 30/05) e deixado O2 3l/min. Foi submetido a laparotomia exploradora em 01/06, sem intercorrências com infestação maciça por ascaris e necrose intestinal, sendo realizada enterectomia -60cm- e ileostomia. Recebeu CHAD novamente no pós operatório (Hb 7,6), precisou de dopa e vários pushes de soro fisiológico. Em 03/06 apresentou RG com coloração fecalóide e eliminou ascaris VO. Fez RX sem evidência de obstrução, tracionada sonda de entérica para gástrica com melhora do RG. Foi extubado, apresentou estridor que cedeu com NBZ com adrenalina e dexta. Apresentou bradicardia. Suspenso fentanil, permaneceu bradicárdico, necessitou atropina. Avaliado pela cardiologia sem particularidades. Após alta da UTI, recebeu óleo mineral e permaneceu eliminando ascaris. Após retirada do óleo mineral foi iniciada dieta, com boa aceitação, e albendazol. Recebe alta em bom estado geral, aceitando dieta VO, SV estáveis, apresentando hiperemia ao redor da ileostomia. Retorno com CIPED.

Source: Hospital de Clínicas de Porto Alegre

Paciente procedente do Hospital de Alvorada onde estava internado desde 28/05 por broncoespasmo e disfunção respiratória, recebendo ampi-sulbactam e beta 2 agonista de 2/2h, com boa evolução. Encaminhado ao HCPA em 01/06 por ter iniciado dia 31/05 com distensão e dor abdominal importantes, acompanhado de vômitos fecalóides, sendo iniciado metro e genta. Havia recebido albendazol durante a internação, sem eliminar vermes. Avaliado pela CIPED: provável obstrução intestinal por ascaris. Na chegada a UTIP feito push soro fisiológico, CHAD (Hb 9 do dia 30/05) e deixado O2 3l/min. Foi submetido a laparotomia exploradora em 01/06, sem intercorrências com infestação maciça por ascaris e necrose intestinal, sendo realizada enterectomia -60cm- e ileostomia. Recebeu CHAD novamente no pós operatório (Hb 7,6), precisou de dopa e vários pushes de soro fisiológico. Em 03/06 apresentou RG com coloração fecalóide e eliminou ascaris VO. Fez RX sem evidência de obstrução, tracionada sonda de entérica para gástrica com melhora do RG. Foi extubado, apresentou estridor que cedeu com NBZ com adrenalina e dexta. Apresentou bradicardia. Suspenso fentanil, permaneceu bradicárdico, necessitou atropina. Avaliado pela cardiologia sem particularidades. Após alta da UTI, recebeu óleo mineral e permaneceu eliminando ascaris. Após retirada do óleo mineral foi iniciado dieta, com boa aceitação, e albendazol. Recebe alta em bom estado geral, aceitando dieta VO, SV estáveis, apresentando hiperemia ao redor da ileostomia. Retorno com CIPED.

Paciente procedente do Hospital de Alvorada onde estava internado desde 28/05 por broncoespasmo e disfunção respiratória, recebendo ampicilina e beta 2 agonista de 2/2h, com boa evolução. Encaminhado à HCPA em 01/06 por ter iniciado dia 31/05 com distensão e dor abdominal importantes, acompanhada de vômitos fecalóides, sendo iniciado metronidazol e gentamicina. Havia recebido albendazol durante a internação, sem eliminar vermes. Avaliado pela CIPED: provável obstrução intestinal por ascaris. Na chegada a UTIP feito push sorológico, CHAD (Hb 9 do dia 30/05) e deixado O2 3l/min. Foi submetido a laparotomia exploradora em 01/06 sem intercorrências com infestação maciça por ascaris e necrose intestinal, sendo realizada enterectomia -60cm- e ileostomia. Recebeu CHAD novamente após operatório (Hb 7,6), precisou de dopa e vários pushes de soro fisiológico. Em 03/06 apresentou RG com coloração fecalóide e eliminou ascaris VO. Fez RX sem evidências de obstrução, traçada sonda de entérica para gástrica com melhora do RG. Foi extubado, apresentou estridor, e cedeu com NBZ com adrenalina e dexa. Apresentou bradicardia. Suspensão de entânico, permaneceu bradicárdico, necessitou atropina. Avaliado pela cardiologia sem particularidades. Após alta da UTI, recebeu óleo mineral e permaneceu eliminando ascaris. Após retirada do óleo mineral, foi iniciado dieta, com boa aceitação, e albendazol. Recebe alta em bom estado geral, aceitando dieta VO, SV estáveis, apresentando hiperemia ao redor da ileostomia. Retorno com CIPED.



Introduction

Methods

Evaluation

Outlook

Summary

Conclusions

- Successful Applied Research in Cross-Language Document Retrieval
 - reaches more than **92%** of an English-English baseline on heterogeneous document collections, in average
- Morphosaurus incorporates six languages:
 - German, English, Portuguese, Spanish, French, Swedish
- Use for Terminology mapping
 - MeSH, ICD, SNOMED CT
- In use in international project consortia and increasingly in commercial systems

Contact

Stefan Schulz

Medical Informatics Department

University Hospital Center Freiburg, Germany

stschulz@uni-freiburg.de

Kornél Markó

Averbis GmbH, Freiburg, Germany

marko@averbis.de

Píndaro Secco Cancian

Paraná Catholic University, Curitiba, Brazil