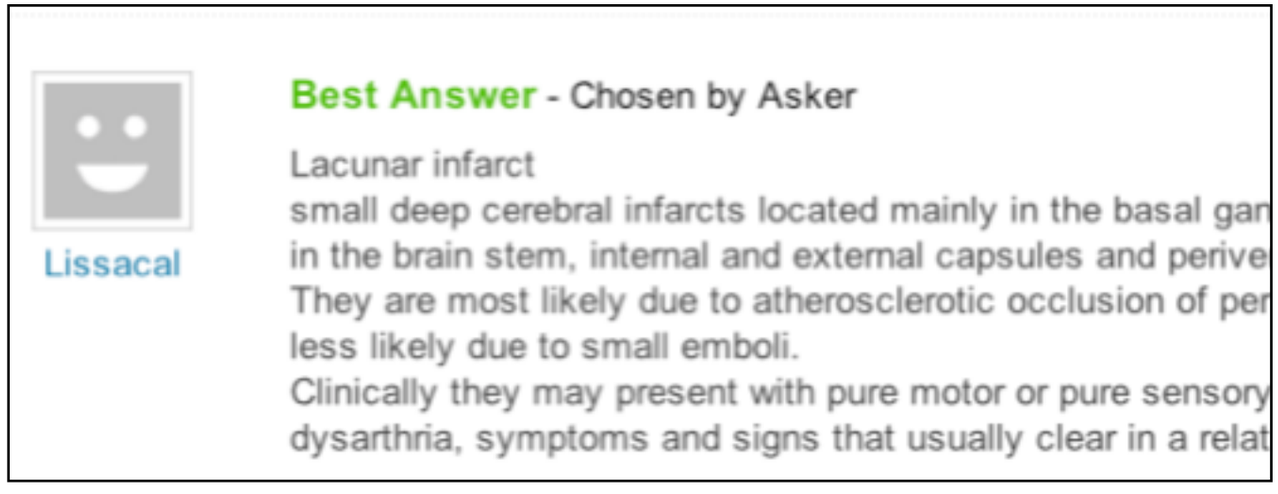
Generating Links to Background Knowledge:

A Case Study Using Narrative Radiology Reports

Jiyin He[1], Maarten de Rijke[2], Merlijn Sevenster[3],
Rob van Ommering[3], Yuechen Qian[3]

[1] CWI; [2] University of Amsterdam; [3] Philips Research

Medical content on the Web



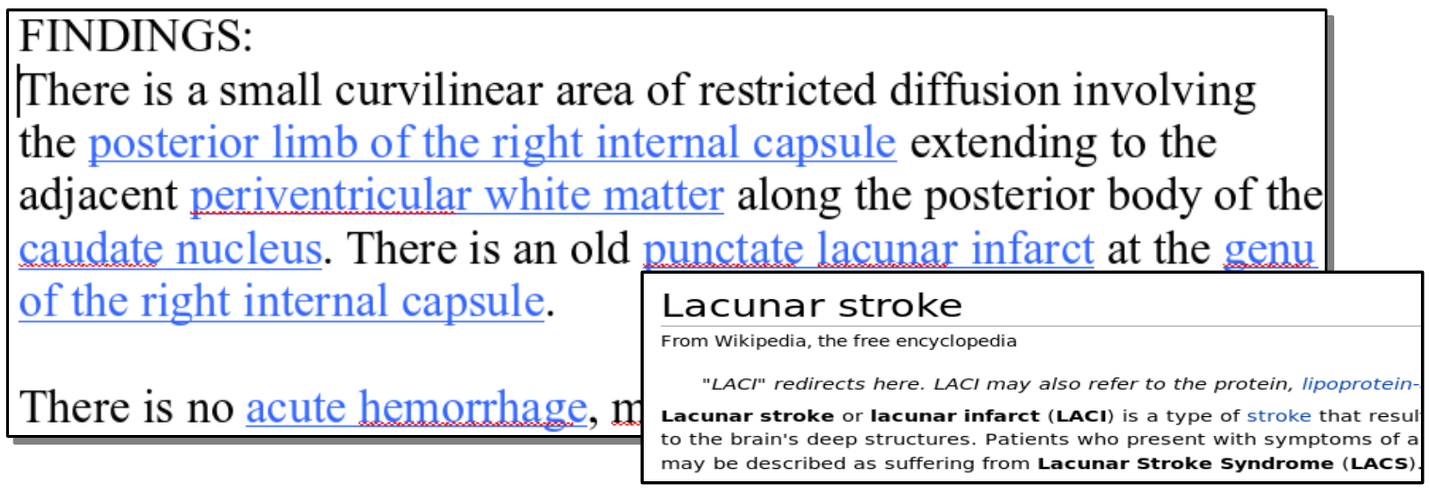
MeSH

Medscape REFERENCE

Medcyclopaedia™ Standard edition

ICD-9

Automatically generate explanatory links to background resources



- In a piece of text, identify terms or phrases that need explanation or background information - **Anchor detection**
  - E.g., medical terminology
- Link it to an item in a knowledge base that provides explanation or background information - **Target finding**
  - E.g., Wikipedia page, ICD descriptions

A case study

- Narrative neuroradiology reports
  - Gives narrative descriptions of the radiologist's findings, diagnoses and recommendations for followup actions
- Wikipedia as background knowledge resource
  - Much work has been done in automatic link generation with Wikipedia in general domain
  - Rich interlinking structure provides valuable training data
  - Covers many medical thesauri and ontologies, e.g., MeSH, ICD-9, ICD-10

A solved problem?

- State-of-the art linking systems
  - E.g., Wikify! (Mihalcea and Csomai, 2007), Wikipedia Miner (Milne and Witten 2008)
  - Exploit Wikipedia link structure
    - Domain independent
- How do they perform in generating links for medical content?
  - An empirical evaluation of existing linking systems on a manually annotated test collection

Two state-of-the-art linking systems

- **Wikify!**
  - Step 1 - Anchor detection:
    - *Keyphraseness* score - the more often a phrase occurs in WP as an anchor text, the more likely it will be used as an anchor text again.
  - Step 2 - Target finding:
    - Lesk algorithm - Measuring the similarity between the context of an anchor text and the target page
    - Machine learning based approach
- **Wikipedia Miner**
  - Step - 1: For each phrase in the current text, finding candidate target pages by measuring the relatedness of a WP page and the context of the phrase
  - Step - 2: Classification to determine the target page for a phrase
  - Step - 3: Classification on anchor - target pairs for anchor detection

Test collection

- 860 anonymized narrative neuroradiology reports
  - 29, 256 anchor - target pairs; 6,440 unique links
  - Anchors are body locations, findings and diagnosis
- Annotated by 3 medical informatics specialists
  - Stage 1: Manually select anchor texts
  - Stage 2: Search for target pages with Wikipedia search engine
    - If no direct matched Wikipedia page was found, a more general concept that reasonably covers the topic was sought
    - If no such page was found, no target was assigned
- Disagreements were resolved through communication (~5% cases)

Experimental setup

- System setup
  - Re-implemented Wikify! ; two versions for target finding - Lesk and machine learning based approach
    - Use Wikipedia miner as a blackbox
- Evaluation metrics: precision, recall and F-measure
- Evaluation on
  - anchor detection
  - target finding - only on correctly identified anchors
  - and overall performance

Results

System	Anchor detection			Target finding			Overall		
	P	R	F	P	R	F	P	R	F
Wikify! (Lesk)	0.35	0.16	0.22	0.4	0.4	0.4	0.14	0.07	0.09
Wikify! (ML)	0.35	0.16	0.22	0.69	0.69	0.69	0.25	0.12	0.16
WM	0.35	0.36	0.36	0.84	0.84	0.84	0.29	0.3	0.3

- Generally not satisfactory
  - only 30% of the links were correctly identified
- Low performance for anchor detection
- Relatively OK performance for target finding

Some observations

- Two properties of the medical anchor texts
  - Regular syntactic structure
    - 70% are noun phrases, where 38 % are single nouns, 32% are nouns with one or more modifiers
    - Can be useful features for anchor detection
  - Complicated semantic structure
    - e.g. "acute cerebral and cerebellar infarction"
    - May cause problems: Wikipedia concepts are usually short and with less complicated structure

	Occurrences in WP links	Coverage	Example
Exact match	923	14.3	"brain" (Report) & "brain" (WP)
Partial match	1,038	16.1	"infarction" (Report) & "cerebellar infarction"(WP)
Sub-exact match	5,257	81.6	"acute cerebral infarction" (Report) & "cerebral infarction" (WP)

Link generation revisited

- The observed structural mismatching between the medical anchor texts and Wikipedia anchor texts causes problems
- Both state-of-the-art systems highly rely on the existing Wikipedia links
- e.g., keyphraseness equals to 0 when a phrase does not occur in WP anchors

CWI

Our approach part 1: anchor detection

- Exploiting the syntactic regularity of medical anchor texts
- A sequential labeling problem: annotate each word of a report with one of the following labels:
  - Begin-of-anchor (BOA); In-anchor (IA); End-of-anchor (EA); Outside-anchor (OA); Single-word-anchor (SWA)
- Conditional random field models (CRFs) with syntactic features
  - The word itself, its POS tag, its syntactic chunk tag

Our approach part II: target candidate identification

- Exploiting existing Wikipedia links with a sub-anchor based approach
  - For a given anchor a, we decompose it into a set of sub-sequences S_a

white matter disease- {white, matter, disease, white matter, matter disease, white matter disease}

- For each sub-anchor s_i, we retrieve top 10 Wikipedia pages as candidates c based on their *target probability:*

The more often a page is linked to a phrase, the more likely it should be linked to it again.

$$p(c_i|s) = \frac{|L_{s,c_i}|}{\sum_{j=1}^{n} |L_{s,c_j}|}$$

CWI

Our approach part III: target detection

- A classification problem: classify each anchor-candidate pair (a, c) as "link" or "non-link"
- Three types of features
  - Title matching - Whether a sub-anchor matches the title of the candidate page; weighted by the similarity of the sub-anchor to the original anchor

$$tm(s,c) = f_{tm}(s,c)\frac{len(s)}{len(a)}, \quad f_{tm}(s,c) = \begin{cases} 1 & \text{if } s \text{ equals title of } c \\ 0 & \text{otherwise.} \end{cases}$$

  - Language model comparison - how likely is the candidate page about neuroradiology?

$$\begin{aligned} \textit{LM-LLR}(T) &= \log\left(\frac{p(T|\theta_R)}{p(T|\theta_W)}\right) \\ &= \sum_{i=1}^{n} \log p(t_i|\theta_R) - \sum_{i=1}^{n} \log p(t_i|\theta_W) \end{aligned}$$

  - Target probability
    - Pre-calculated at candidate identification stage
    - Aggregate from sub-anchor level to anchor level: Max, Min, Avg

Experiment setup

- 3-fold cross-validation
- Classifiers for target detection:
  - SVM, NB and Random Forest*
- A post-processing step for target detection
  - If all candidates are classified as “non-link”, the one with the lowest confidence score is chosen
  - If multiple candidates are classified as “link”, the one with the highest confidence score is chosen

Evaluation

- Anchor detection

System	P	R	F
LiRa	**0.9**	**0.8**	**0.85**
Wikify!	0.35	0.16	0.22
WM	0.35	0.36	0.36

Results of anchor detection
LiRa: system using our proposed approach

Evaluation

- Target finding

System	P	R	F
LiRa	**0.8**	**0.8**	**0.8**
Wikify! (Lesk)	0.4	0.4	0.4
Wkify! (ML)	0.69	0.69	0.69

Results of target finding for anchors identified by Wikify!

System	P	R	F
LiRa	**0.68**	**0.68**	**0.68**
Wikify! (Lesk)	0.13	0.13	0.13
Wikify! (ML)	0.26	0.26	0.26

Results of target finding for annotated anchors

System	P	R	F
LiRa	**0.89**	**0.89**	**0.89**
WM	0.84	0.84	0.84

Results of target finding for annotated anchors

Evaluation

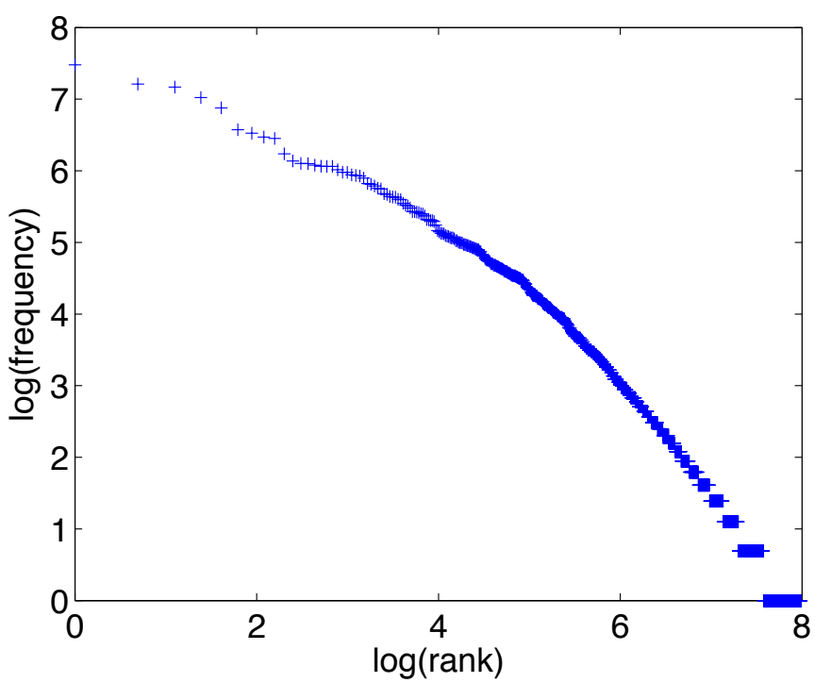
- Overall performance

System	P	R	F
LiRa	**0.65**	**0.58**	**0.61**
Wikify! (Lesk)	0.14	0.07	0.09
Wikify! (ML)	0.25	0.12	0.16
WM	0.29	0.3	0.3

Impact of anchor frequencies

- Some anchors occur more frequent than others
  - Frequent anchors are likely to be general concepts
  - More likely to occur in Wikipedia
  - Large amount of infrequent anchors, few frequent anchors

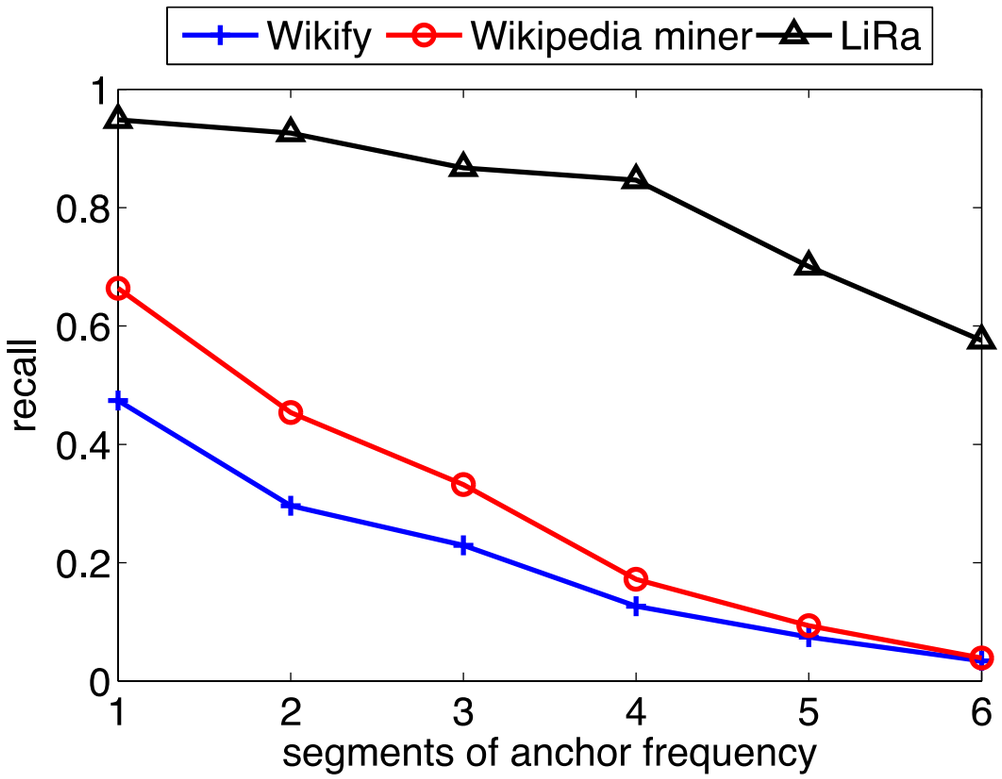
Top 5	Bottom 5
mass	vestibular nerves
brain	Virchow-Robin space
meningioma	Warthin's tumor
frontal	Wegner's granulomatosis
white matter	xanthogranulomas

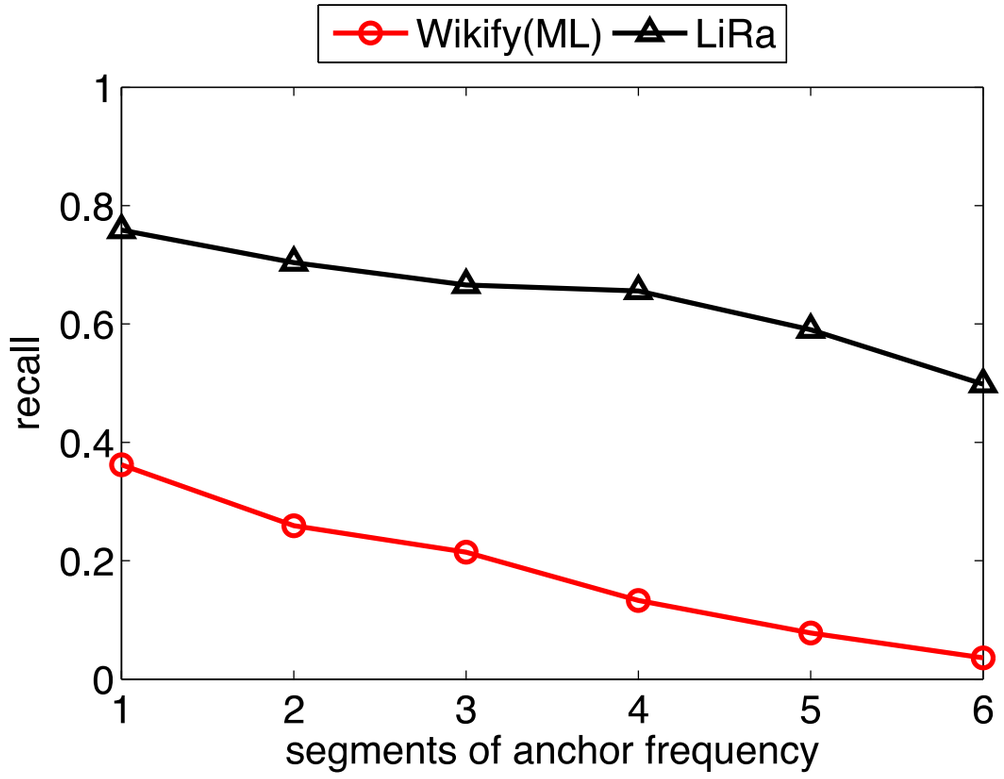


Impact of anchor frequencies

- How does this influence the performance of linking systems?

Group	1	2	3	4	5	6
Freq. range	>100	51-100	11-50	6-10	2-5	1
#Anchors	116	108	527	482	1,399	2,149





Conclusions

- Existing link generation systems trained on general domain corpora do not provide a satisfactory solution to linking radiology reports
- Structural mismatch between medical phrases and Wikipedia concepts is a major problem
- Our proposed approach was shown to be effective
- Frequent anchor texts tend to be "easier" than anchor texts with a low frequency

Conclusions

- Existing link generation systems trained on general domain corpora do not provide a satisfactory solution to linking radiology reports
- Structural mismatch between medical phrases and Wikipedia is a major problem
- Our proposed approach was shown to be effective
- Frequent anchor texts tend to be “easier” than anchor texts with a low frequency

Questions?