Automated Summarisation for Evidence Based Medicine

Diego Mollá

Centre for Language Technology, Macquarie University

HAIL, 22 March 2012



590

Contents

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Our Corpus for Summarisation

Applications 0000 000000 000000

About Us: Research Group on Natural Language Processing of Medical Texts

http://web.science.mq.edu.au/~diego/medicalnlp/

Active Members

Diego Mollá Senior lecturer at Macquarie University.

Cécile Paris Senior principal research scientist at CSIRO ICT Centre.

Abeed Sarker PhD student at Macquarie University.

Sara Faisal Shash Masters student.

Past Members

María Elena Santiago-Martínez Research programmer.

Patrick Davis-Desmond Masters student.

Andreea Tutos Masters student.



Contents

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Our Corpus for Summarisation

Evidence Based Medicine



http://laikaspoetnik.wordpress.com/2009/04/04/evidence-based-medicine-the-facebook-of-medicine/



EBM and Natural Language Processing

Frame Patient Scenario into a Clinical Question Librarian Centred Systematically retrieve best evidence available THE NEW YORKER ABINA COULD BE Critically appraise evidence ANYTHING Apply results to patient Evaluate decision making WAY TOO GENERAL PRACTITIONER MACQUARIE

http://hlwiki.slais.ubc.ca/index.php?title=Five_steps_of_EBM

PICO for Asking the Right Question

	1	2	3	4	
	Patient or Problem	Intervention (a cause, prognostic factor, treatment, etc.)	Comparison Intervention (if necessary)	Outcomes	
Tips for Building	Starting with your patient, ask "How would I describe a group of patients similar to mine?" Balance precision with brevity.	Ask "Which main intervention am I considering?"	Ask "What is the main alternative to compare with the intervention?" Again, be specific	Ask "What can I hope to accom- plish?", or "What could this exposure really affect?" Again, be specific	
Example	"In patients with heart failure from dilated cardiomyopathy who are in sinus rhythm"	"would adding anticoagulation with warfarin to standard heart failure therapy"	"when compared with standard therapy alone"	"lead to lower mortality or morbidity from thromboembolism. Is this enough to be worth the increased risk of bleeding?"	



Where to search for external evidence?

- 1. Evidence-based Summaries (Systematic Reviews):
 - EBM Online (http://ebm.bmj.com).
 - UptoDate (http://www.uptodate.com).
 - ► The Cochrane Library (http://www.thecochranelibrary.com/).
 - ▶ ...



Where to search for external evidence?

- 1. Evidence-based Summaries (Systematic Reviews):
 - EBM Online (http://ebm.bmj.com).
 - UptoDate (http://www.uptodate.com).
 - ► The Cochrane Library (http://www.thecochranelibrary.com/).
 - ▶ ...
- 2. Search the Medical Literature:
 - ► E.g. PubMed (http://www.ncbi.nlm.nih.gov/pubmed/).



Searching Cochrane

Home About us Cochrane Reviews News & Events Training Multimedia Contact Community	<u>ary</u>
	90
Home > Home	
Use the tabs below to filter your search results:	
Cochrane Reviews Colloquia abstracts Methodology Newsletters Policy Manual Registered titles Whole site	
Enter your keywords:	
sleep apnea	
Search Showing results 1 to 10 of 12	
Anti-inflammatory drugs for the treatment of obstructive sleep Skip to main content Cochrane Summaries Anti-inflammatory drugs for the treatment of obstructive sleep apnea in children summaries.cotrane.org/CD00707uttive-sleep-apnea-in-children	
Pharmacotherapy for hypertension in adults with obstructive Skip to main content Cochrane Summaries Cochrane Summaries beta Pharmacotherapy for hypertension in adults with obstructive sleep apnea summaries.cochrane.org/cD0766swith-obstructive-sleep-apnea	
Adenotonsillectomy for obstructive sleep apnoea in children [The Cochrane Library Adenotonsillectomy for obstructive sleep apnoea in children. Have your say! Your UNIVERSITY	

EBM Summarisation

Searching PubMed

S NCBI Resources 🕑 How To 🗹	My NCBI Sign In
Public Public iseep apnea US National Indicated Indic	Search Help
Display Settings: © Summary, 20 per page, Sorted by Recently Added Send to: © Results: 1 to 20 of 24722 << Fest. < Prev. Page 1 of 1237 Next.> Lat>>	Filter your results: All (24722) Free Full Text (5987)
Importance and management of chronic sleep apnoea in cardiology, Jaffe LM, Kjekshus J, Gottlieb SS. Eur Heart J. 2012 Mar 16. [Epub ahead of print] PMID: 22472872 PubMid-a saypoiled by publisher]	Review (4058) Manage Filters
Related citations Association of inflammatory biomarkers with steep disorders in hemodialysis patients. Racepti E, Sahraian MA, Heidari R, Bagherzadeh M. Acan Neuro Beigo 2014 Mrc112(14)-56 Epub 2012 (Feb 2. PMUD: 22427289 (PiaMud - In process) Related citations	Related searches central sleep apnea obstructive sleep apnea syndrome sleep apnea heart sleep apnea cardiovascular cleap apnea trainmort
Exploration of the relationship between sleep position and isolated longue base or multilevel surgery in obstructive sleep aprea. van Maanen JP, Ravesloot MJ, Witte BI, Grijseels M, de Vries N. Eur Arch Odminolaropat, 2021 Zur Zu Digba barea do rphet PMID: 22472104 [PubMed - as supplied by publisher] Detailer drafters	Titles with your search terms Snoring and obstructive sleep apnea. [Med Clin North Am. 2010] Obstructive sleep apnea: diagnosis,
Attention in children with obstructive sleep apnoea: An event-related potentials study, A Barnes NE, Gozal D, Molfese DL, Sleep Med. 2012 Mo 14. [Epub ahead of print]	epidemiology, and ecol [Respir Care. 2010] Cardiovascular consequences of sleep apnea. [Clin Chest Med. 2010] See more



MACQUARIE UNIVERSITY

Searching the Trip Database

Home About Login Register I	Labs		frip	
Q sleep apnea Search Tips		Search Advanced Search	History Translate	
FILTER SEARCH		SEARCH RESULTS	ASSOCIATED RESULTS	
Order By: Date Relevance (Showing all results - Only show it	new)	Select All Choose Your Action	MEDLINE ARTICLES	Pub Ced
EVIDENCE All Secondary Evidence Evidence Based Synopses Systematic Reviews	5,797 738 169 274	1. Continuous positive airway pressure delivery interfaces for obstructive sleep apnoea COCHRANE DATABASE OF SYSTEMATIC REVIEWS 2011	Therapy Etiology Diagnosis Prognosis Systematic Reviews	887 1,252 539 1,034 201
Guidelines Aus. & NZ Canada UK	29 22 63	CONTRACT OF COMPANY TOTAL TO A CONTRACT OF COMPANY TO A CONTRACT OF COMPANY TO A CONTRACT OF COMPANY TO A CONTRACT OF CON	CLINICAL TRIALSClinic 706 trials	alTrials.go
USA other Clinical Q&A Core primary research Extended primary research e Textbooks	168 13 60 136 3,408 855	CPROME Developing Wold? Related Community Preview DOI Anti-inflammatory medications for obstructive sleep agnea in children CocReve DarAbase Co Systematic Reserves 2001 CPDICME Developing Wold? Related Community Preview DOI	BNF RESULTS An error has occurred	BNF
Patient Decision Aids Patient Information More News MEDICAL IMAGES	3 276 83 238 15	4. An integrated health-economic analysis of diagnostic and therapeutic strategies in the treatment of moderate-to-severe obstructive sleep apnea Heal Economic Reduktion NortAnce. 2012 CPDIOME Developing Work? Belated Constance Perview	RELATED ARTICLES	see related.



Appraising the Evidence

The SORT Taxonomy

- Level A Consistent and good-quality patient-oriented evidence.
- Level B Inconsistent or limited-quality patient-oriented evidence.
- Level C Consensus, usual practise, opinion, disease-oriented evidence, or case series for studies of diagnosis, treatment, prevention, or screening.



Levels of Evidence

Study quality	Diagnosis	Treatment / prevention / screening	Prognosis
Level 1: good-quality patient-oriented evidence	Validated clinical decision rule; SR/meta-analysis of high-quality studies; high- quality diagnostic cohort study	SR/meta-analysis of RCTs with consistent findings; high-quality individual RCT; all-or-none study	SR/meta-analysis of good- quality cohort studies; prospective cohort study with good follow-up
Level 2: limited-quality patient-oriented evidence	Unvalidated clinical decision rule; SR/meta- analysis of lower-quality studies or studies with inconsistent findings; lower-quality diagnostic cohort study or diagnostic case-control study	SR/meta-analysis of lower- quality clinical trials or of studies with inconsistent findings; lower-quality clin- ical trial; cohort study; case-control study	SR/meta-analysis of lower- quality cohort studies or with inconsistent results; retrospective cohort study or prospective cohort study with poor follow-up; case- control study; case series
Level 3: other evidence	Consensus guidelines, extrapo disease-oriented evidence (in series for studies of diagnosis	plations from bench research, use termediate or physiologic outco , treatment, prevention, or scre	ual practice, opinion, omes only), or case ening



Where can NLP Help?

- Questions:
 - Help to formulate answerable questions.
 - Question analysis and classification.





Where can NLP Help?

- Questions:
 - Help to formulate answerable questions.
 - Question analysis and classification.
- ► Search:
 - Retrieve and rank relevant literature.
 - Extract the evidence-based information.
 - Summarise the results.





Where can NLP Help?

- Questions:
 - Help to formulate answerable questions.
 - Question analysis and classification.
- ► Search:
 - Retrieve and rank relevant literature.
 - Extract the evidence-based information.
 - Summarise the results.
- Appraisal: Classify the evidence.



Contents

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Where's the Corpus for Summarisation?

Summarisation Systems

- CENTRIFUSER/PERSIVAL: Developed and tested using user feedback (iterative design).
- SemRep: Evaluation based on human judgement.
- ► Demner-Fushman & Lin: ROUGE on original paper abstracts.
- ► Fiszman: Factoid-based evaluation.



Where's the Corpus for Summarisation?

Summarisation Systems

- CENTRIFUSER/PERSIVAL: Developed and tested using user feedback (iterative design).
- SemRep: Evaluation based on human judgement.
- Demner-Fushman & Lin: ROUGE on original paper abstracts.
- Fiszman: Factoid-based evaluation.

Corpora

- ► Several corpora of questions/answers available.
- Answers lack explicit pointers to primary literature.
- Medical doctors want to know the primary sources.

Contents

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus

How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Our Corpus for Summarisation 000

Journal of Family Practice's "Clinical Inquiries"

Which treatments work best for hemorrhoids?

Evidence-based answer

Excision is the most effective treatment for thrombosed external hemorrhoids (strength of recommendation [SOR]: B, retrospective studies). For prolapsed internal hemorrhoids, the best definitive treatment

Evidence summary

the dentate line and become acutely painful with thrombosis. They can cause perianal pruritus and excoriation ed nonsurgically found improved pain because of interference with perianal relief with a combination of topical hygiene, Internal hemorrhoids become nifedipine 0.3% and lidocaine 1.5% symptomatic when they bleed or pro- compared with lidocaine alone. The apse (TABLE).

For thrombosed external hemorrhoids, surgery works best

Few studies have evaluated the best treatment for thrombosed external Many studies have evaluated the best 53% in the group receiving fiber.8 hemorrhoids. A retrospective study treatment for prolapsed hemorrhoids. of 231 patients treated conservatively A Cochrane systematic review of 12 When surgical hemorrhoidectomy or surgically found that the 48.5% RCTs that compared conventional is recommended of patients treated surgically had a hemorrhoidectomy with stapled hem- The American Society of Colon and lower recurrence rate than the conser- orrhoidectomy in patients with grades Rectal Surgeons recommends adequate vative group (number needed to treat I to III hemorrhoids found a lower fluid and fiber intake for all patients [NNT]=2 for recurrence at mean fol- rate of recurrence (follow-up ranged with symptomatic hemorrhoids, For low-up of 7.6 months) and earlier reso- from 6 to 39 months) in patients who grade I to III hemorrhoids, the society lution of symptoms (average 3.9 days had conventional hemorrhoidectomy states that handing is usually most efcompared with 24 days for conserva- (NNT=14).4 Conventional hemorrhoid- fective. When office treatments fail, the tive treatment).1

340 patients who underwent outpa- continence, tient excision of thrombosed external



is traditional hemorrhoidectomy (SOR: A, systematic reviews). Of nonoperative techniques, rubber band ligation produces the lowest rate of recurrence (SOR: A, systematic reviews).

ported a low recurrence rate of 6.5% External hemorrhoids originate below at a mean follow-up of 17.3 months.²

A prospective, randomized controlled trial (RCT) of 98 patients treat-NNT for complete pain relief at 7 days was 3.3

Conventional hemorrhoidectomy beats stanling

Another retrospective analysis of in decreased bleeding and decreased in- rhoidectomy (SOR: B).

A second systematic review of 25 thrombosed hemorrhoids less than 72 hemorrhoids under local anesthesia re- studies, including some that were of hours old and expectant treatment with

lower quality, showed a higher recurrence rate at 1 year with stapled hemorrhoidectomy than with conventional surgery,5

Nonoperative techniques? Consider rubber band ligation

A systematic review of 3 poorquality trials comparing rubber band ligation with excisional hemorrhoidectomy in patients with grade III hemorrhoids found that excisional hemorrhoidectomy produced better long-term symptom control but more immediate postoperative complications of anal stenosis and hemorrhoids that present early. Surgihemorrhage.6 Rubber band ligation had techniques of sclerotherapy and infra- grade III and IV hemorrhoids.10 red coagulation.3

Fiber supplements help relieve symptoms

A Cochrane systematic review of 7 RCTs enrolling a total of 378 patients with grade I to III hemorrhoids evaluated the effect of fiber supplements on pain, itching, and bleeding. Persistent hemorrhoid symptoms decreased by

ectomy showed a nonsignificant trend society recommends surgical hemor-

The society recommends excision of

Classification of sympto internal hemorrhoid

GRADE	DESCRIPTION
1	Hemorrhoids do not protr
	Hemorrhoids protrude wit reduce spontaneously
ш	Hemorrhoids protrude an by hand
IV	Hemorrhoids are perman
Source: Madoff R	D. et al. Gestroentero/ooy 2004.10

cal hemorrhoidectomy should be rethe lowest recurrence rate at 12 months served for when conservative treatment compared with the other nonoperative fails and for patients with symptomatic

- 1. Gmenspon J. Williams SB. Young HA, et al. Thrombased external hemomoids: outcome after conservative or surgical management. Dis Colon Rectum: 2004:47:1493-1498.
- Jonnen J. Bach S, Stubinger SH, et al. Excision of thrombosed external hemorrhoids under local anesthesia: a retrospective evaluation of 340 patients, Dis Colon Rectum, 2003;46:1226-1231
- 3. Perrotti P. Antropoli C. Molino D. et al. Conservative treatment of acute thrombosed external hemorthoids with topical nifedipine. Dis Colon Rectum
- 4. Javaraman S. Colouhoun PH. Malthaner RA. Stapled versus conventional surpery for bemonhoids Cochrane Database Syst Rev. 2006;(4):CD005383.
- 5. Tiandra JJ, Chan MK, Systematic review on the procedure for prolapse and hemorrhoids (stapled hemorrhoidopexy). Dis Colon Rectum
- Systematic review of randomized trials comparing rubber band ligation with excisional haemorrhoid ectomy, Br J Sura, 2005;92:1481-1487.
- 7. Johanson JF, Rimm A. Optimal nonsurgical treatment of hemorrhoids: a comparative analysis of infrared conquistion, rubber band ligation and injection scientherapy. Am J Gastroenterol 1992:87:1600-1606.
- Alonso-Coello P. Guyatt G, Heels-Ansdell D, et al Laxatives for the treatment of hemorrhoids. Cocrivane Database Syst Rev. 2005(4):CD004649.

Applications 0000 000000 000000

The XML Contents I

```
<record id ="7843" >
<url>http://www.jfponline.com/Pages.asp?AID=7843&amp;issue=September_2009&amp;UID=</url>
<question>Which treatments work best for hemorrhoids?</question>
<answer>
  <snip id="1">
    <sniptext>Excision is the most effective treatment for thrombosed
external hemorrhoids.</sniptext>
    <sor type="B">retrospective studies</sor>
    <long id ="1_1">
      longtext>A retrospective study of 231 patients treated
      conservatively or surgically found that the 48.5% of patients
      treated surgically had a lower recurrence rate than the
      conservative group (number needed to treat [NNT]=2 for
      recurrence at mean follow-up of 7.6 months) and earlier
      resolution of symptoms (average 3.9 days compared with 24 days
      for conservative treatment). </longtext>
      <ref id="15486746" abstract="Abstracts/15486746.xml">Greenspon
      J. Williams SB. Young HA .et al. Thrombosed external
      hemorrhoids: outcome after conservative or surgical
      management. Dis Colon Rectum. 2004; 47: 1493-1498.</ref>
    </long>
    <long id ="1_2">
      <longtext>A retrospective analysis of 340 patients who underwent
      outpatient excision of thrombosed external hemorrhoids under
      local anesthesia reported a low recurrence rate of 6.5% at a
```



The XML Contents II

```
mean follow-up of 17.3 months.</longtext>
      <ref id ="12972967" abstract="Abstracts/12972967.xml">Jongen J,
      Bach S, Stubinger SH , et al. Excision of thrombosed external
      hemorrhoids under local anesthesia: a retrospective evaluation
      of 340 patients. Dis Colon Rectum, 2003: 46: 1226-1231.</ref>
    </long>
   <long id ="1_3">
      <longtext>A prospective . randomized controlled trial (RCT) of 98
      patients treated nonsurgically found improved pain relief with a
      combination of topical nifedipine 0.3% and lidocaine 1.5% compared
      with lidocaine alone. The NNT for complete pain relief at 7 days was
      3.</longtext>
      <ref id="11289288" abstract="Abstracts/11289288.xml">Perrotti P.
      Antropoli C, Molino D , et al. Conservative treatment of acute
      thrombosed external hemorrhoids with topical nifedipine. Dis
      Colon Rectum, 2001: 44: 405-409.</ref>
    </long>
  </snip>
</answer>
</record>
```



Components of the Corpus

Question direct extract from the source.

Answer split from the source and manually checked.

Evidence extracted from the source.

Additional text manually extracted from the source and massaged.

References PMID looked up in PubMed (automatic and manual procedure).



Contents

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Annotation of Text Justifications

Goal

- Identify the text justifications.
- Align the text justifications with the answer parts.

Method

LINIV/FRSIT

- ► Three annotators (members of the research group).
- Annotation tool contains pre-zoned text:
 - answer summary;
 - body text;
 - recommendations;
 - references.

Annotators need to copy and paste (and massage) the text.

Annotation Tool I

		-					
JFP Corpus Annotation Tool							
	Pag						
	URL http://www.ifponline.com/Pages.asp? AID=1080&issue=January_2002&UID=						
	Title What is the most effective treatment for tinea pedis athlete's foot?						
	Aut	hors	Tsveti Markova, MD				
Help - How	to Annotate						
ANSWER	s						
SNIP ID			SNIP TEXT	SOR TYPE	SOR BA	SES REFERENCES	
1	Topical therapy is effective for tinea pedis. Topical terbinafine has a 70% cure rate, is available over the counter DTC, and requirs only it to 2 weeks of therapy. Two other OTC (points, linkfihler and microarde), regard = 2 to 4 weeks to active the None						
1_1				1	1		
+Long							
2	The most effective tre for 2 weeks 94% clinic approved for this indic hyperkerototic soles, t	atment cal cure cation. I severe	for tinea pedis is oral terbinafine 250 mg twice a day rate. However, oral terbinafine is expensive and not Oral therapy may be required for patients with disease, topical therapy failure, chronic infection or	В	based or small randomi d	ze v	
2_1							
+Long							
+Snip	Snip						



Annotation Tool II

SUMMARY

The Cochrane Database of Systemic Reviews, reported 72 placebo-controlled trials of topical agents that yielded the following cure rates: undecenoic acid. 72%; allylamines terbinafine, naftifine, butenafine, 70%; tolnaftate, 64%; azoles miconazole, clotrimazole, ketoconazole, econazole, oxiconazole, 47%, A meta-analysis of 11 RCTs suggests that allylamines are slightly more effective than azoles. (REF:1,2).

Orally administered antifungal agents are expensive and can have systemic side effects. Griseofulvin and ketoconazole are approved for oral therapy, but product labels clearly state that they should be used only after topical agents have failed. Griseofulvin has been used for more than 30 years, is well tolerated, and efficacious in treating dermatomycoses in the range of 60%. Ketoconazole's cure rate is similar, but its use in cutaneous infections is limited by multiple drug interactions and serious side effects. Three placebo-controlled RCTs of itraconazole of varying doses and duration of treatment suggested favorable clinical cure of moccasin-type tinea pedis 51%-85%. The most effective itraconazole regimen was 200 mg twice Jaily for 1 week. In a large double-blind multicenter study of all forms of tinea pedis. De Keyser et al compared 2 weeks of terbinafine at 250 mg/day to 2 weeks of itraconazole at 100 mg/day. After 8 weeks they found terbinafine superior to itraconazole for clinical cure 94.1% vs 72.4%. In a single multicenter open study the cure rate for fluconazole 150 mg was 77% when used once weekly for 3 weeks. (REF:3.4),

RECOMMENDATIONS

American Academy of Dermatology Guidelines recommend topical therapy for initial treatment of tinea pedis. Oral therapy may be required to treat patients with hyperkeratotic soles, disabling or extensive disease, topical therapy failure, chronic infection, or immunosuppression. Surgical therapy is not indicated. (REF:5).

REFERENCES

MAC UNIV

ID	PUBMED	CORRECT PUBMED	SOR TYPE	PUB TYPE	CITATION
1	<u>19040832</u>	[]			Crawford F, Hart R, Bel-Syer S, Togerson D, Young P, Russell I . Cochrane Review. In: The Cochrane Library, Issue 3, 2001. Oxford: Update Software.
2	20685791				Hart R, Sally E, Bell-Syer S, Crawford F, Togerson D, Young P, Russell I . BMJ 1999; 319: 79-82.
3	20967420				Pierard G, Arrese J, Pierrard-Franchimont C . Drugs 1996; 52: 209.
4	None				De Keyser P, De Backer M, Massart DL, Westelnick KJ . Br J Dermatol 1994; 130: 22-5.
٥	20947733 RIE				Drake LA, Dinehart SM, Farmer ER ,et al. J Am Acad Dermatol 1996; 34: 282-6.
RSI	TY 🎵	//			EBM Summarisation

Annotating Answer Justifications

Conventions for text massaging

- 1. Remove/edit connecting phrases.
- 2. Remove irrelevant introductory text.
- 3. If a paragraph has several references, attempt to split the paragraph.
 - May need to massage the text of resulting splits.
- 4. If a paragraph has no references, attempt to merge with previous or next paragraph.



Finding PubMed IDs

Method

- 1. Split the reference text into sentences.
- 2. Remove author and pagination text:
 - Use simple regexps.
- 3. Perform a sequence of searches with all combinations of sentences.



Example I

Collins NC . Is ice right? Does cryotherapy improve outcome for acute soft tissue injury? Emerg Med J. 2008; 25: 65-68.

- Collins NC .
- ► Is ice right?
- Does cryotherapy improve outcome for acute soft tissue injury
- Emerg Med J. 2008; 25: 65-68.



Example II

list	search	ID	title	match %
1, 2, 3	Is ice right? Does cryotherapy improve outcome for acute soft tissue injury? Emerg Med J	18212134	Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?	92
1, 2	Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?	18212134	Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?	100
1, 3	Is ice right? Emerg Med J	18212134	Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?	39
2, 3	Does cryotherapy improve out- come for acute soft tissue injury? Emerg Med J	18212134	Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?	82
1	Is ice right?	None	None	0
2	Does cryotherapy improve out- come for acute soft tissue injury?	15496998	Does Cryotherapy Improve Out- comes With Soft Tissue Injury? 78	
3	Emerg Med J	None	None	0



Using Amazon Mechanical Turk I

Mechanics

- AMT was used to find the correct IDs.
- An AMT hit had 10 references:
 - ▶ 2 known references for checking quality of annotation.
- Each hit was assigned to 5 Turkers.
- There was a preliminary training session.



Using Amazon Mechanical Turk II

Approving and rejecting hits

Reject hit if there are two or more "bad" IDs, i.e. one of:

- A known ID is wrong.
- The ID is invalid:
 - Not found in PubMed;
 - No title is returned.
- The title of the ID does not match the title of our reference:
 - threshold: 50% match.
- The ID does not agree with majority.



Applications 0000 000000 000000

Using Amazon Mechanical Turk III

Checking validity for final annotation

- Majority wins automatically except when:
 - majority is a "bad" ID;
 - majority is the "nf" ID;
 - ▶ the other two are agreeing ("full house").
- Manual check is done in all other cases.


Contents

Evidence Based Medicine

Our Corpus for Summarisation

Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Corpus Statistics

Size

- ▶ 456 questions ("records").
- ▶ 1,396 answers ("snips").
- ▶ 3,036 text explanations ("longs").
- 3,705 references:
 - ▶ 2,908 unique references.
 - 2,657 XML abstracts from PubMed.



Answers per Question



Snips per Record



Answer justifications per answer



References per answer justification







References per question



EBM Summarisation

References per Record





Evidence Grade



number of SORs



References



number of references



Contents

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Contents

Evidence Based Medicine

Our Corpus for Summarisation

Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses

Single-document Summarisation Evidence Grading



Evidence-based Summarisation

Single Document Summarisation

Input: Question, reference. Target: Text explanation.



Evidence-based Summarisation

Single Document Summarisation

Input: Question, reference.

Target: Text explanation.

Multi-document Summarisation

Input: Question, group of relevant references. Target: Answer parts (optional: plus text explanation).



Appraisal, Clustering

Text Classification for Appraisal

Input: Group of references. Target: Evidence-based grade.



Appraisal, Clustering

Text Classification for Appraisal

Input: Group of references.

Target: Evidence-based grade.

Clustering

Input: Question, group of relevant references. Target: Cluster groupings (optional: plus answer parts).



Retrieval?

Possible task

Input: Question. Target: List of references.



Retrieval?

Possible task

Input: Question. Target: List of references.

However...

- Some of the references are old.
- The references are likely not exhaustive.



Contents

Evidence Based Medicine

Our Corpus for Summarisation

Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



Input, Output

Input

- Question.
- Document Abstract.

Output

- Extractive summary that answers the question.
- Target summary is the annotated evidence text ("long").
- Evaluated using ROUGE-L with Stemming.



Baselines

plain Return the last *n* sentences.

keywords Return the last *n* sentences that share any non-stop words with the question.

umls Return the last *n* sentences that share any UMLS concepts with the question.

System	F	Conf Interval	
baseline plain	0.193	[0.190–0.196]	 •
baseline keywords	0.195	[0.192–0.198]	
baseline umls	0.194	[0.190–0.197]	o



Preselect sentences and then:

	Abstract		
	Section 1	S1.1 S1.2	
	Section 2	S2.1	
	Section 3	S3.1 S3.2	
	Section 4	S4.1 S4.2	
	Section 5	S5.1 S5.2	
MA UN	IVERSITY		

Summary

Preselect sentences and then:

1. Use PubMed's section tags (background, conclusions, methods, objective, results).

	Abstract		
	Background	S1.1 S1.2	
	Methods	S2.1	
	Results	S3.1 S3.2	
	Conclusions	S4.1 S4.2	
	Conclusions	S5.1 S5.2	
MA	cquarie 🎦 🍿 👘		
UN	IVERSITY		ED

M.

Summary

Μ

Using the Abstract Structure

Preselect sentences and then:

- 1. Use PubMed's section tags (background, conclusions, methods, objective, results).
- 2. Select the first *n* sentences of the last "conclusions" section.

	Abstract		
	Background	S1.1 S1.2	
	Methods	S2.1	
	Results	\$3.1 \$3.2	
	Conclusions	S4.1 S4.2	
	Conclusions	S5.1 S5.2	
MA	cquarie 🎦 🎁		
UN	IVERSITY		EDM

Summary	
S5.1 S5.2	

Preselect sentences and then:

- 1. Use PubMed's section tags (background, conclusions, methods, objective, results).
- 2. Select the first *n* sentences of the last "conclusions" section
- 3. If we have less than *n* sentences, fill from the first sentences of the previous "conclusions" section, and so on until all "conclusions" sections are used up.

	Abstract		
	Background	S1.1 S1.2	
	Methods	S2.1	
	Results	\$3.1 \$3.2	
	Conclusions	S4.1 S4.2	
	Conclusions	S5.1 S5.2	
MA	CQUARIE		
UΝ	IVERSITY		F

Su	m	ma	ry				
	_	~-	~	~ .	_	~ .	

Preselect sentences and then:

- 1. Use PubMed's section tags (background, conclusions, methods, objective, results).
- 2. Select the first *n* sentences of the last "conclusions" section.
- 3. If we have less than *n* sentences, fill from the first sentences of the previous "conclusions" section, and so on until all "conclusions" sections are used up.
- 4. If we have less than *n* sentences, fill from the "results" sections.

Abstract			Summary	
Background Methods Results Conclusions	\$1.1 \$1.2 \$2.1 \$3.1 \$3.2 \$4.1 \$4.2 \$5.1 \$5.2		S5.1 S5.2 S	4.1 S4.2 S3.1
		EBM S	ummarisation	Diego M

Preselect sentences and then:

- 1. Use PubMed's section tags (background, conclusions, methods, objective, results).
- 2. Select the first *n* sentences of the last "conclusions" section.
- 3. If we have less than *n* sentences, fill from the first sentences of the previous "conclusions" section, and so on until all "conclusions" sections are used up.
- 4. If we have less than *n* sentences, fill from the "results" sections.
- 5. If we still have less than n sentences, fill from the "methods" sections.

	Abstract			Summary	
N/ A	Background Methods Results Conclusions Conclusions	\$1.1 \$1.2 \$2.1 \$3.1 \$3.2 \$4.1 \$4.2 \$5.1 \$5.2		S5.1 S5.2 S4.2	L S4.2 S3.1
UN	IVERSITY		EDM		Diana

Preselect sentences and then:

- 1. Use PubMed's section tags (background, conclusions, methods, objective, results).
- 2. Select the first *n* sentences of the last "conclusions" section.
- 3. If we have less than *n* sentences, fill from the first sentences of the previous "conclusions" section, and so on until all "conclusions" sections are used up.
- 4. If we have less than *n* sentences, fill from the "results" sections.
- 5. If we still have less than n sentences, fill from the "methods" sections.
- 6. If the abstract has no structure, return the last n sentences.

	Abstract				Summary
	Background Methods Results Conclusions Conclusions	\$1.1 \$1.2 \$2.1 \$3.1 \$3.2 \$4.1 \$4.2 \$5.1 \$5.2			S5.1 S5.2 S4.1
MA UN			F	BM Summa	risation

S4.2 S3.1

Results

The F is calculated using ROUGE-L with stemming.

System	F	Conf Interval	
baseline plain baseline keywords baseline umls	0.193 0.195 0.194	[0.190-0.196] [0.192-0.198] [0.190-0.197]	
structure plain structure keywords structure umls	0.196 0.193 0.192	[0.193–0.199] [0.190–0.197] [0.189–0.195]	0 0



ROUGE-L with Stemming for All 3-Sentence Subsets I

Process

MAC

- 1. Compute the ROUGE-L of all 3-sentence subsets in each abstract.
- 2. Find the decile boundaries in each abstract.
- 3. Find the distribution of decile boundaries.

	0	1	2	3	4	5
Mean Std Dev	0.094 0.060	0.136 0.062	0.153 0.065	0.164 0.067	0.176 0.070	0.188 0.073
		6	7	8	9	10
	Mean	0.200	0.213	0.229	0.249	0.299
UARIE	Std Dev	0.076	0.081	0.087	0.094	0.112

Our Corpus for Summarisation 0000 0000000 0000000 Applications

ROUGE-L with Stemming for All 3-Sentence Subsets II



ollá 52/60

Contents

Evidence Based Medicine

Our Corpus for Summarisation

Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading



ALTA 2011 Shared Task

The ALTA Shared Tasks

- Competitions where all participants are evaluated on the same data.
- ▶ The ALTA 2011 shared task was based on evidence grading.

The Data

- Clusters of abstracts.
- The SOR grade of each cluster.



Data Sample

Fragment

41711 B 10553790 15265350 53581 C 12804123 16026213 14627885 53583 B 15213586 52401 A 15329425 9058342 11279767



Words as Features

Abstract *n*-grams

- Generated *n*-grams (n = 1, 2, 3, 4) for each of the abstracts.
- Replaced specific medical concepts with generic 'sem_type' tags using UMLS.
- Stemmed, lowercased, stop words removed.

Title *n*-grams

- Generated *n*-grams (n = 1, 2) for each title.
- Processed in the same way as abstract n-grams.



Publication Types as Features I

Distribution of publication types in a different corpus.



- Meta-analysis
- Crossectional/Crossover Study
- Expert Opinion

- Systematic Review
- 🖾 Other Clinical Trial
- Case Series/Report
- Consensus/Usual Practice
- Randomized Controlled Trial ② Review 巴 Other Study ■ Unknown



Publication Types as Features II

Publication types

- ► Rule-based classifier to detect publication types.
- Simple regular expressions that identify major publication types.
- Used the publication types marked up by PubMed when available.
- If an article has several possible publication types, choose the one with highest quality.



М

Cascaded Classification

Process: Cascaded SVMs

- 1. Default class: B.
- 2. SVMs with abstract *n*-grams to identify A and C.
- 3. SVMs with publication types to identify A and C.
- 4. SVMs with title *n*-grams to identify A and C.

Results	5				
	Method	Accuracy	Confidence Interv	als	
	Majority (B) Cascaded SVMs	48.63% 62.84%	41.5 - 55.83		
		EBM Summarisation		Diego Mollá	5
Questions?

Evidence Based Medicine

Our Corpus for Summarisation Structure of our Corpus How we Created the Corpus Statistics

Applications

Possible Uses Single-document Summarisation Evidence Grading

Further Information

http://web.science.mq.edu.au/~diego/medicalnlp/

