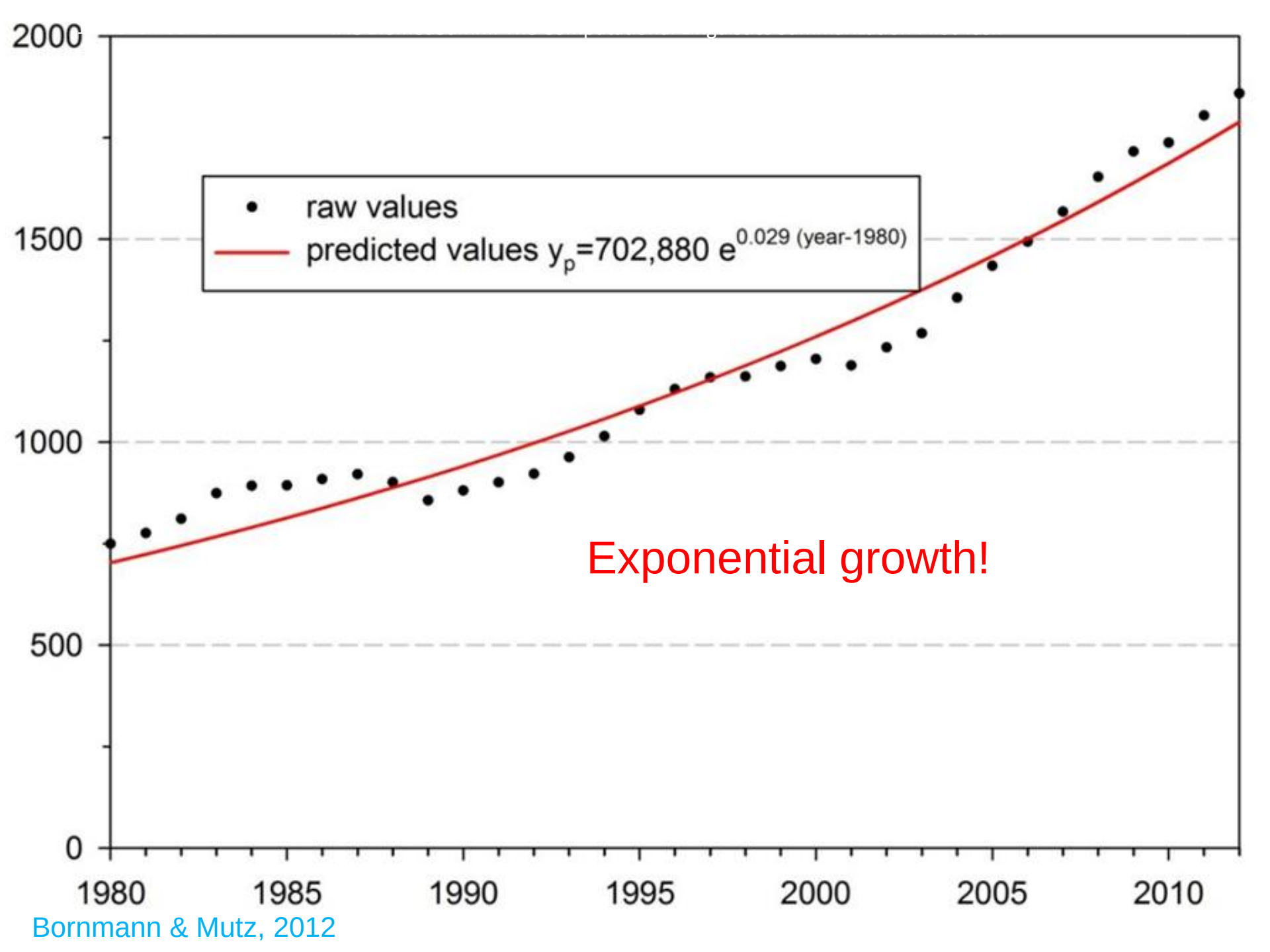
The Computational Linguistics Summarization Pilot Task @ TAC 2014

Kokil Jaidka*, Muthu Kumar Chandrasekaran, Rahul Jha, Christopher Jones, Min-Yen Kan, Ankur Khanna, Diego Molla-Aliod, Dragomir R. Radev, Francesco Ronzano, Horacio Saggion

* Nanyang Technological University, Singapore
koki0001@e.ntu.edu.sg

Scientific Document Summarization

Photo Credits Dennis Jarvis @flickr



Bornmann & Mutz, 2012

Outline

- CL-Summ so far
  - Citation based extractive summaries
  - Faceted summaries
  - Automatic literature review
  - ACL corpus
- The CL-Summ Shared Task
- TAC 2015: CL-Summ track
- Acknowledgements

Scientific Document Summarization

- Abstract
  - Authors' own summary
- Citation summary
  - Community creates a summary when citing
- Faceted summary
  - Capture all aspects of a paper

Towards Automated Related Work Summarization

Cong Duy Vu Hoang and **Min-Yen Kan**
Department of Computer Science
School of Computing
National University of Singapore
{hcdvu,kanmy}@comp.nus.edu.sg

Abstract

We introduce the novel problem of automatic related work summarization. Given multiple articles (*e.g.*, conference/journal papers) as input, a related work summarization system creates a topic-biased summary of related work specific to the target paper. Our prototype **Re**lated **Wo**rk **S**ummarization system, **ReWoS**, takes in set of keywords arranged in a hierarchical fashion that describes a target paper's topics, to drive the creation of an extractive summary using two different strategies for locating appropriate sentences for general

Framework for Scientific document summarization

We dissect the full challenge as bringing together work of disparate interests; 1) in finding relevant documents; 2) in identifying the salient aspects of these documents in relation to the current work worth summarizing; and 3) in generating the final topic-biased summary. While it is clear that current NLP technology does not let us build a complete solution for this task, we believe that tackling the individual components will help

Using Citations to Generate Surveys of Scientific Paradigms

Saif Mohammad[†*], Bonnie Dorr[†‡*], Melissa Egan[†‡], Ahmed Hassan[ϕ], Pradeep Muthukrishan[ϕ], Vahed Qazvinian[ϕ], Dragomir Radev[ϕ§], David Zajic[‡◇]

Institute for Advanced Computer Studies[†] and Computer Science[‡], University of Maryland.
Human Language Technology Center of Excellence[*]. Center for Advanced Study of Language[◇].
{saif,bonnie,mkegan,dmzajic}@umiacs.umd.edu

Department of Electrical Engineering and Computer Science[ϕ]
School of Information[§], University of Michigan.
{hassanam,mpradeep,vahed,radev}@umich.edu

Citations 1) select papers and 2) identify salient parts of the cited paper.

first steps in producing an automatically generated, readily consumable, technical survey. Specifically we explore the combination of citation information and summarization techniques. Even though prior work (Teufel et

some of which may be unfamiliar to panelists. Thus, they must learn about a new discipline "on the fly" in order to relate their own expertise to the proposal.

Our goal is to effectively serve these needs by combining two currently available technologies: (1) bibliometric lexical link mining that exploits the structure of citations and relations among citations; and (2) summarization techniques that exploit the content of the material in both the citing and cited papers.

It is generally agreed upon that manually written

Scientific Document Summarization
Citation based extractive summaries

Scope of Citation

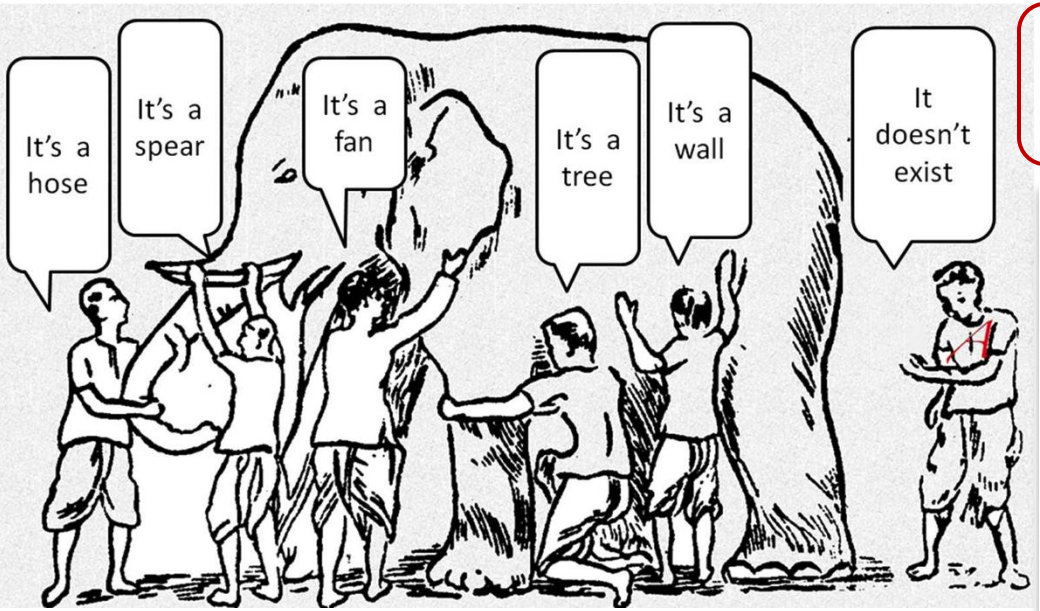
- Qazvinian, V., and Radev, D. R. “Identifying non-explicit citing sentences for citation-based summarization” (ACL, 2010)

- Abu-Jbara, Amjad, and Dragomir Radev. “Reference scope identification in citing sentences.” (ACL, 2012)

Citation summary

Blind Men and Elephants: What Do Citation Summaries Tell Us About a Research Article?

Aaron Elkiss, Siwei Shen, Anthony Fader, Güneş Erkan, David States, and Dragomir Radev
University of Michigan, Ann Arbor MI 48109



The old Asian legend about the blind men and the elephant comes to mind when looking at how different authors of scientific papers describe a piece of related prior work. It turns out that different citations to the same paper often focus on different aspects of that paper and that neither provides a full description of its full set of contributions. In this article, we will describe our investigation of this phenomenon. We studied *cita-*

Citations are like celebrity quotes taken out of context

[Image credits Ken Ammi](#) @flickr

Deconstructing Human Literature Reviews – A Framework for Multi-Document Summarization

Kokil Jaidka, Christopher S.G. Khoo, Jin-Cheon Na
Division of Information Studies
Wee Kim Wee School of Communication and Information
Nanyang Technological University, Singapore
[kokil, chriskhoo]@pmail.ntu.edu.sg, tjcna@ntu.edu.sg

Scientific summaries are written to fulfill information & argumentative functions

- Identify new research questions
- Define the proposed research contributions
- Build the justification for the current work
- Situate the work in the research literature
- Reinterpret and critique previous results

These rhetorical characteristics of literature reviews make it a challenging research problem in automatic multi-document summarization – not only should the summarizer identify salient information, but it should also synthesize the summary in a way that achieves certain argumentative purposes. The problem of summarization in context was first identified by Sparck Jones and Endres-Niggemeyer (1995) and

Scientific Document Summarization
Citation-based extractive summaries

Scope of Citation

- Qazvinian, V., and Radev, D. R. “Identifying non-explicit citing sentences for citation-based summarization” (ACL, 2010)
- Abu-Jbara, Amjad, and Dragomir Radev. “Reference scope identification in citing sentences.” (ACL, 2012)

Coherence

- Abu-Jbara, Amjad, and Dragomir Radev. “Coherent citation-based summarization of scientific papers.” (ACL 2011)

A B S T R A C T

Objective: Obesity and disturbed eating behaviors are both associated with low self-este
body images. Th
evolution of mer
were evaluated
Methods: In tota
program lasting
after 18 weeks, a
addressing healt
to body schema
Results: Obese p
quality of life (p
image ($p < 0.001$).
Conclusion: DTW allowed obese patients to reset both their somatic and psychic cons
body image.
Practice implications: Patients are usually reluctant to practice physical activity. Dance
not only body image, but also psycho-social aspects of their personality.

Faceted summaries are structured abstracts.

Common in domains such as Medicine, Biomedical, Bioinformatics

Summarising Scientific Articles — Experiments with Relevance and Rhetorical Status

Simone Teufel*
Cambridge University

Marc Moens*
Rhetorical Systems
and University of Edinburgh

In this paper we propose a strategy for the summarisation of scientific articles which concentrates on the rhetorical status of statements in the article: material for summaries is selected in such a way that summaries can highlight the new contribution of the source paper and situate it with respect to earlier work.

Argumentative zones as facets demarcate new contributions of a paper from background work.

In summary,

- Community concurs that a citation based summary of a scientific document is important to create
- Citing papers cite different points of the same reference paper
- Assigning facets to these citances may help create coherent summaries

Outline

- CL-Summ ground work done so far
- The CL-Summ corpus, task evaluation
  - Highlights
  - Annotation
  - Evaluation results
- TAC 2015: CL-Summ track
- Acknowledgements

CL-Summ Pilot: highlights

First corpus in the computational linguistics community incorporating prior research on citation based summaries

- 10 teams registered
- 3 teams participated in the evaluation
- 2 teams submitted their systems' performance
- 1 more proposed algorithms to solve the tasks

All Fields | Search... | Search

Login
Bookmarks | History

Oct 2014: The Third CIPS-SIGHAN Joint Conference on Chinese Language Processing has now been added to the Anthology. The Proceedings of Traitement of Automatique des Langues Naturelles 2014 and its five associated workshops are now available on the Anthology.

Welcome to the ACL Anthology

The ACL Anthology currently hosts 33846 papers on the study of computational linguistics and natural language processing. Subscribe to the mailing list to receive announcements and updates to the Anthology.

ACL Events	Present - 2010					2009 - 2000										1999 - 1990										1989 - 1974									
CL	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79-74
TACL	14	13																																		
ACL	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79
EACL	14		12			09			06			03				99		97		95		93		91		89		87		85		83				
NAACL		13	12		10	09		07	06		04	03		01	00																				
*SEMEVAL	14	13	12		10			07			04			01			98																			
ANLP															00			97			94		92				88					83				
EMNLP	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96																	
CONLL	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97																		
WS	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96	95	94	93		91	90											
SIGs	ANN | BIOMED | DAT | DIAL | FSM | GEN | HAN | HUM | LEX | MEDIA | MOL | MT | NLL | PARSE | MORPHON | SEM | SEMITIC | SLPAT | WAC																																		

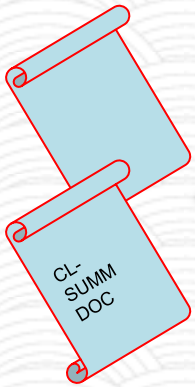
Non-ACL Events	Present - 2010					2009 - 2000										1999 - 1990										1989 - 1974									
COLING	14		12		10		08		06		04		02		00		98		96		94		92		90		88		86	82	80			73	69	67	65
HLT		13	12		10	09	08	07	06	05	04	03		01							94	93	92	91	90	89			86								
IJCNLP		13		11		09	08			05																										
LREC	14		12		10		08		06		04		02		00																					
PACLIC		13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98		96	95																
ROCLING		13	12	11	10	09	08	07	06	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	89	88									
TINLAP																												87				78	75			
ALTA		13	12	11	10	09	08	07	06	05	04	03																								
RANLP		13		11		09																														

CL training Corpus

- 10 reference papers or topics randomly sampled from the ACL live anthology
- Up to 10 citing papers per reference paper including those outside ACL live anthology
- Annotated corpus publicly available

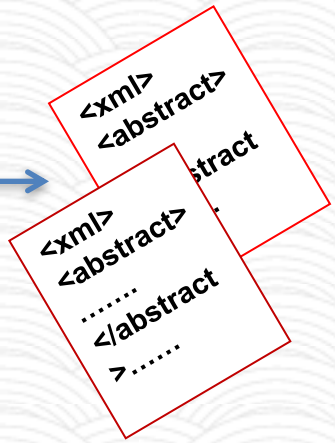
https://github.com/WING-NUS/scisumm-corpus/

Annotation Pipeline

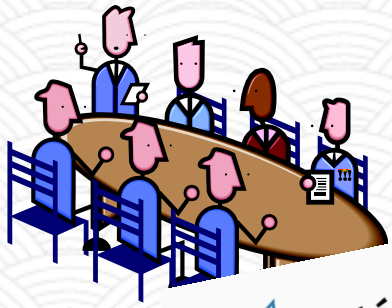


OCR & Section Parse

ParsCit's SectLabel module



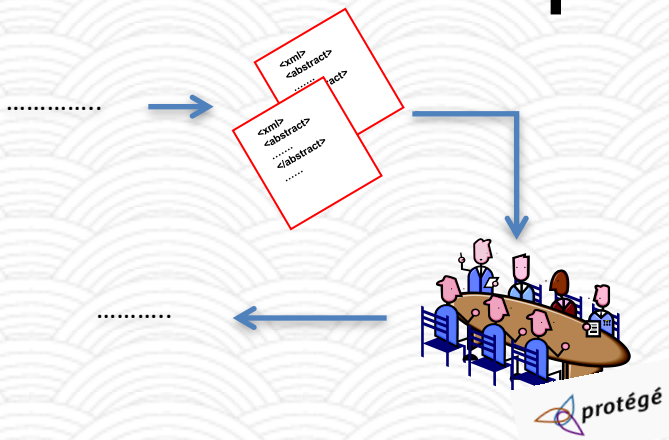
Annotation!



protégé

Post Processing to BiomedSumm format

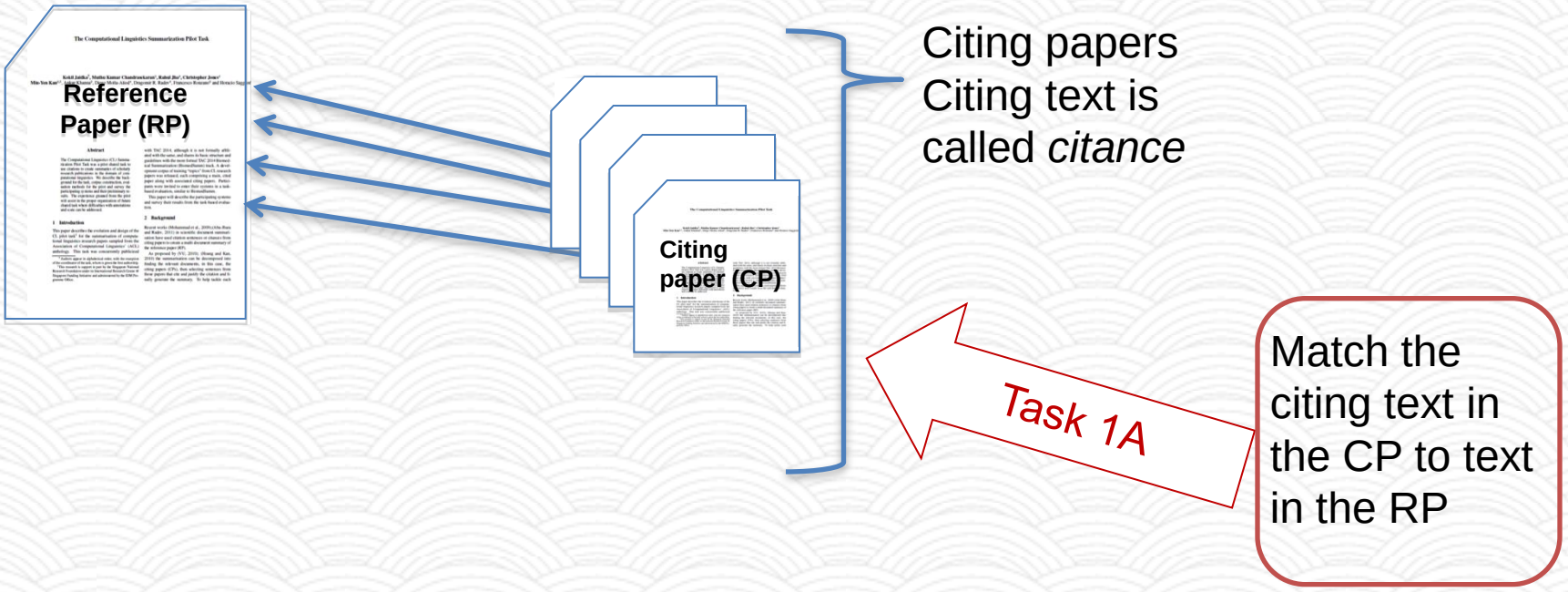
Annotating the SciSumm corpus



- 3 annotators
- Released data has one gold standard annotation per topic or reference paper
- Discourse facet has a minor change from Biomedsumm's categories

Tasks

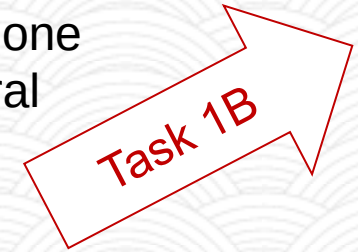
Task 1A: Identify the text span in the RP which corresponds to the *citances* from the CP.

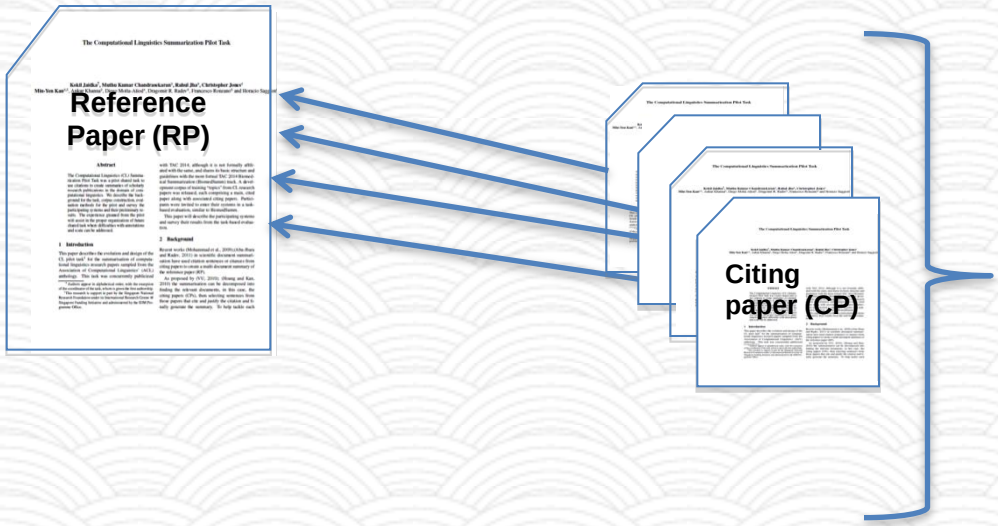


Tasks

Task 1B: Identify the discourse facet for every cited text span from a predefined set of facets.

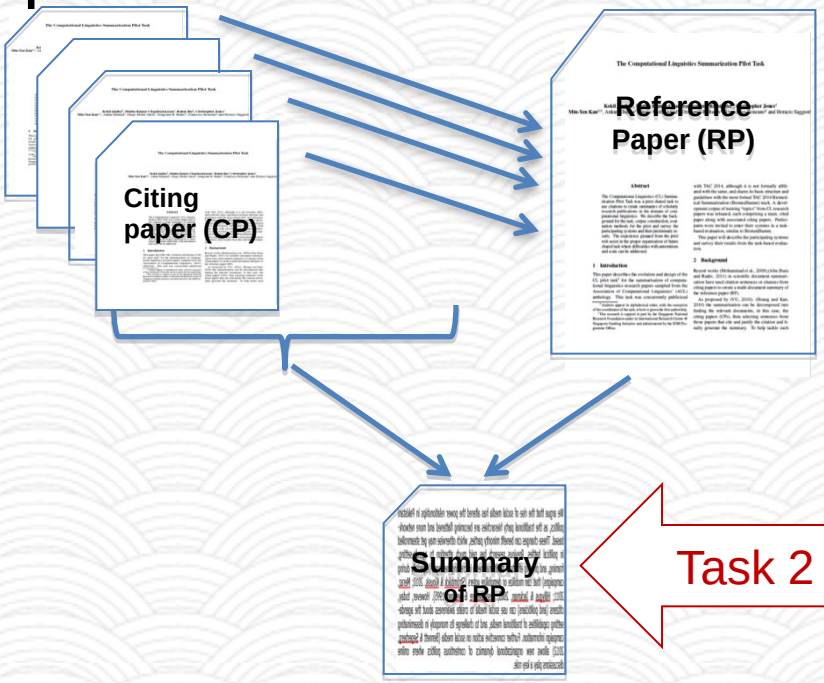
Classify the cited text in RP into one of several facets





Tasks

Task 2: Generate a faceted summary of up to 250 words, of the reference paper, using itself and the citing papers.



Use citances and the RP to create a summary

Evaluation

Small corpus: 10 fold cross validated evaluation over the 10 documents

- Task 1A scored by ROUGE-L metric
- Task 1B scored by classification metrics: Precision, Recall and F_1
- Task 2 also scored by ROUGE-L metric

Results – Task 1A

MQ			Clair_UMich		
Precision	Recall	F_1	Precision	Recall	F_1
0.212	0.335	0.223	0.444	0.574	0.487

- MQ was unsupervised while Clair_Umich was supervised
- Challenging classification problem: Task seeks to map each citation sentence with a few out of 100s of potential matches in the Reference paper (RP)
- Lexical, semantic and structural similarities between citances and RP sentences somewhat help

Results – Task 1A

Paper ID	MQ	Clair_UMich
C90_2039	0.235	0.635
C94_2154	0.288	0.536
E03_1020	0.239	0.478
H05_1115	0.350	0.375
H89_2014	0.332	0.546
J00_3003	0.196	0.559
J98_2005	0.101	0.344
N01_1011	0.221	0.498
P98_1081	0.200	0.367
X96_1048	0.248	0.535

Large deviation in scores, across topics, from both systems

Results – Task 2

Paper ID	MQ (using Task 1A MMR)
C90_2039	0.293
C94_2154	0.120
E03_1020	0.196
H05_1115	0.321
H89_2014	0.320
J00_3003	0.367
J98_2005	0.233
N01_1011	0.284
P98_1081	0.206
Average	**0.260**

ROUGE-L scores here measure overlap over the abstract since we did not have human summaries

Low scores could be due to deviation between summary of citances and the abstract of the paper

Errors – Task 1A

Citing text: "The line of our argument below follows a proof provided in… for the maximum likelihood estimator based on nite tree distributions."

Clair_UMich

False negative: "We will show that in both cases the estimated probability is tight."

MQ

Target text from RP: "*The work described here also makes use of hidden Markov model.*"

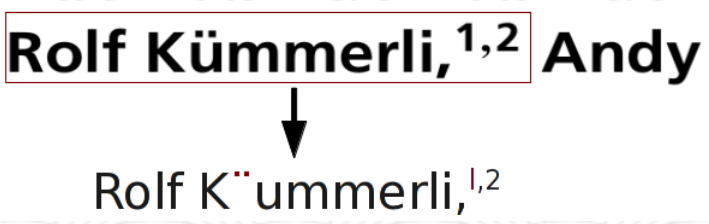
False positive: "The statistical methods can be described in terms of Markov models."

Learning from the Pilot Task

- Offset mismatch between the text file and the XML that annotators used
  - Corpus sentence segmented and sentences assigned a sentence ID
- Problems in post-processing non-contiguous annotated reference spans.
- Character offsets can be miscounted by different parsers
- Handling non-UTF8 characters

Limitations of this corpus

- No gold standard citation based summaries
- OCR errors:



- The use of "..." where text spans are snippets
- Errors in citation/reference offset numbers
- Different text encodings
- Errors in file construction
- Small size of corpus!

Acknowledgements

- NIST and Hoa Dang
- Lucy Vanderwende, MSR,
- Anita de Ward, Elsevier Data Services
- Kevin B. Cohen, Prabha Yadav (U. Colorado, Boulder)
- Horacio Saggion for detailed bug report on the corpus
- Rahul Jha (U. Mich, Ann Arbor)
- All BiomedSumm track participants

Questions? Thank you!