Linear Algebra in Document Summarization

Symposium in Honor of Dianne P. O'Leary
SIAM Conference Applied Linear Algebra 2015

John M. Conroy
IDA Center for Computing Sciences
Bowie, MD
USA

Google

refugee crisis europe

Web | **News** | Videos | Images | Maps | More ▾ | Search tools

About 6,340,000 results (0.35 seconds)

Refugee crisis: Slovenia calls in army to help patrol borders
The Guardian - 6 hours ago
Slovenia on Tuesday called in the army to help it manage **refugees** seeking to reach northern **Europe** before winter, as the small EU state ...

Europe's Refugee Crisis: Slovenia to Deploy Its Army
The Atlantic - Oct 20, 2015

Europe refugee crisis: Slovenia to use army to help control border flow
CBC.ca - Oct 20, 2015

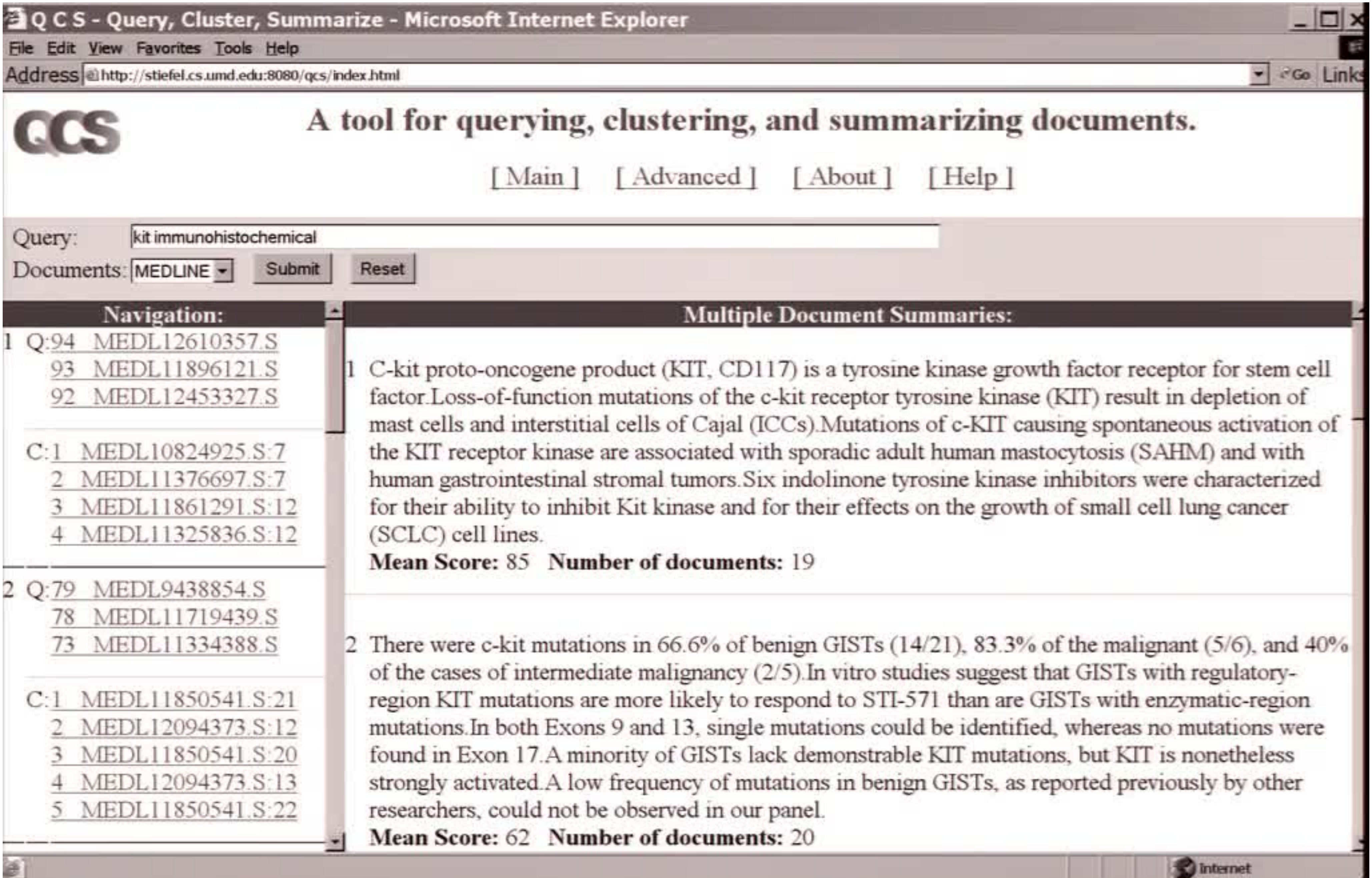
Slovenia calls army to border to handle asylum seeker surge
Opinion - ABC Online - 20 hours ago

Europe refugee crisis: Slovenia may raise border fence against ...
In-Depth - Sydney Morning Herald - Oct 20, 2015

The disintegration of **Europe**
Opinion - Times of Oman - Oct 19, 2015

Sydney M... | The Atlantic | CBC.ca | Internation... | Deutsche ... | Irish Times

Explore in depth (1,783 more articles)

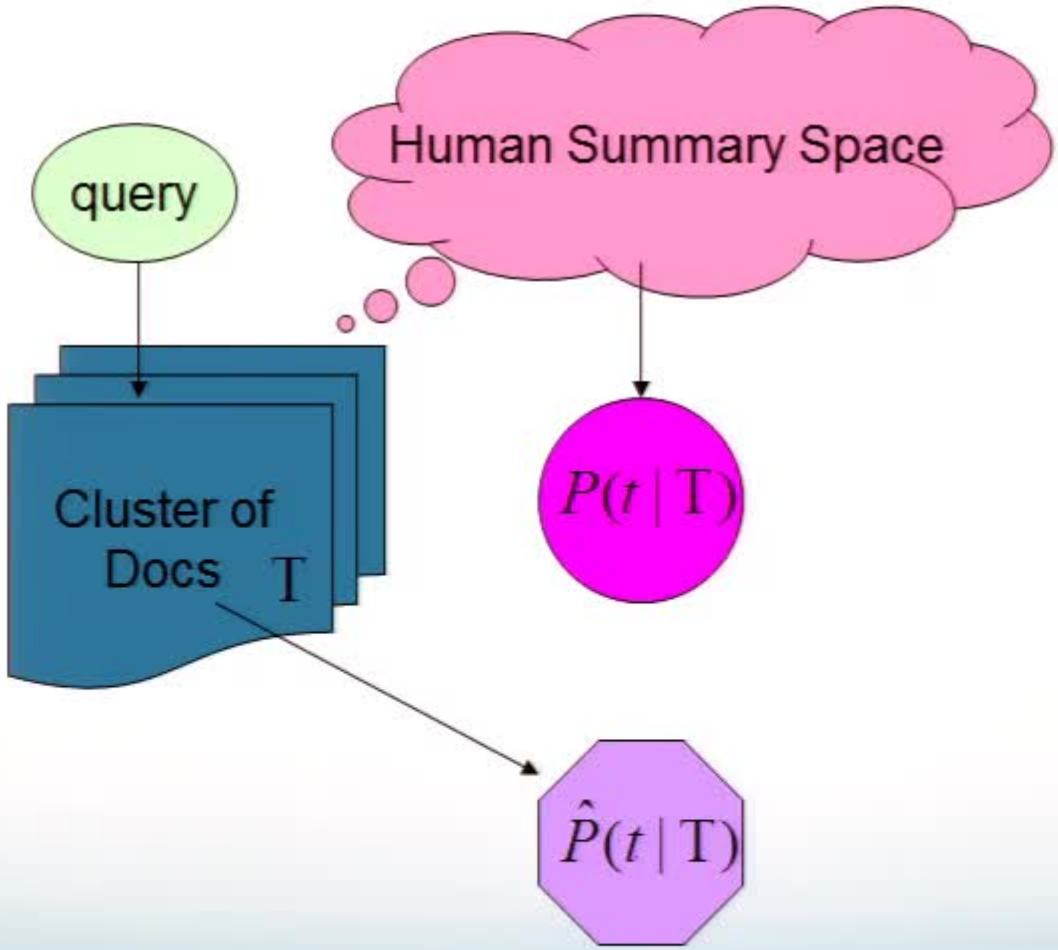


Dunlavy, D.M., O'Leary, D.P., Conroy, J.M. & Schlesinger, J.D.. QCS: A System for Querying, Clustering and Summarizing Documents, Information Processing & Management Vol. 43(6), pp. 1588-1605. 2007.

Summarization Task

- Given a cluster of documents, create a fluent synopsis of specified length using information in the document cluster.
- Model Problem:
  - English newswire documents.
- Other Applications:
  - technical documents, Multi-lingual, email, speech transcripts.
- Can also be used to improve information retrieval.
  T. Sakai and K. Spärk Jones, ``Generic summaries for indexing in information retrieval'', *SIGIR 01, Proceedings of 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2001.

A Model of Human Abstracts



An Approximation of $P(t|\tau)$

- We approximate $P(t|\tau)$ by

$$P_{qs\rho}(t \mid \tau) = \alpha_q q(t) + \alpha_s s(t) + \alpha_\rho \rho(t)$$

$$s(t)[q(t)] = \begin{cases} 1 \text{ if } t \text{ is a signature [query] term} \\ 0 \text{ if } t \text{ is not a signature [query] term} \end{cases}$$

$$\rho(t \mid \tau) = \text{ probability } t \text{ occurs in a sentence considered for selection.}$$

- The score of a sentence is the sum of $P(t)$ taken over its terms divided by its length.
- Conroy, O' Leary, Schlesinger ACL/COLING 2006.

An Approximation of $P(t|\tau)$

- We approximate $P(t|\tau)$ by

$$P_{qs\rho}(t \mid \tau) = \alpha_q q(t) + \alpha_s s(t) + \alpha_\rho \rho(t)$$

$$s(t)[q(t)] = \begin{cases} 1 \text{ if } t \text{ is a signature [query] term} \\ 0 \text{ if } t \text{ is not a signature [query] term} \end{cases}$$

$$\rho(t \mid \tau) = \text{ probability } t \text{ occurs in a sentence considered for selection.}$$

- The score of a sentence is the sum of $P(t)$ taken over its terms divided by its length.
- Conroy, O' Leary, Schlesinger ACL/COLING 2006.

Two Approaches

- Pivoted *QR*.
  - Conroy & O'Leary SIGIR 2001
- A "L1-non-negative" "*QR*," *NR* factorization
  - Non-negative factorization $A \approx NR$.
  - Generally ``pre-condition" matrix by using a truncated SVD to remove noise.
  - Conroy, O' Leary, Schlesinger ACL/COLING 2006.

Pivoted QR

Use scores to select candidate sentences (~6*w*)

- Terms as sentence features
  - Terms: $\{t_1, \ldots, t_m\} \varepsilon \boldsymbol{R}^m$
  - Sentences: $\{s_1, \ldots, s_n\} \varepsilon \boldsymbol{R}^n$
  - Scaling: $\| \boldsymbol{a} \|_2$ =score for sentence.

$$\begin{array}{c|ccc} & s_1 & \cdots & s_n \\ \hline t_1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ t_m & a_{m1} & \cdots & a_{mn} \end{array}$$

- Perform QR with column pivoting, implicit elimination to preserve sparsity.
- Stop criteria: chosen sentences (columns) have ~*w* words.

QR Inspired Non-negative Matrix Factorization

Use scores to select candidate sentences (6*w* words).

$$\begin{array}{c|ccc} & s_1 & \cdots & s_n \\ \hline t_1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ t_m & a_{m1} & \cdots & a_{mn} \end{array}$$

- Terms as sentence features
  - Terms: $\{t_1, \ldots, t_m\} \varepsilon \boldsymbol{R}^m$
  - Sentences: $\{s_1, \ldots, s_n\} \varepsilon \boldsymbol{R}^n$
  - Scaling: $\| \boldsymbol{a} \|_1$ =score for sentence.
- Maintain non-negativity but we lose orthogonality: $A \approx NR$.
- Choose column with maximum 1-norm ($\boldsymbol{a}_j$)
- Subtract components along $\boldsymbol{a}_j$ from remaining columns, but set negatives to 0.
- Stop criteria: chosen sentences (columns) have ~*w* words

Evaluation Methods

- Manual Evaluation
  - Overall Responsiveness (grade for summaries 1-5 scale)
  - Pyramid (a content metric)
- Automatic Evaluation
  - ROUGE-1 (unigrams)
  - ROUGE-2 (bigrams)
  - ROUGE-3 (trigrams)
  - ROUGE-4 (tetra-grams)

Lin, Chin-Yew and E.H. Hovy 2003. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In Proceedings of 2003 Language Technology Conference (HLT-NAACL 2003), Edmonton, Canada, May 27 - June 1, 2003.

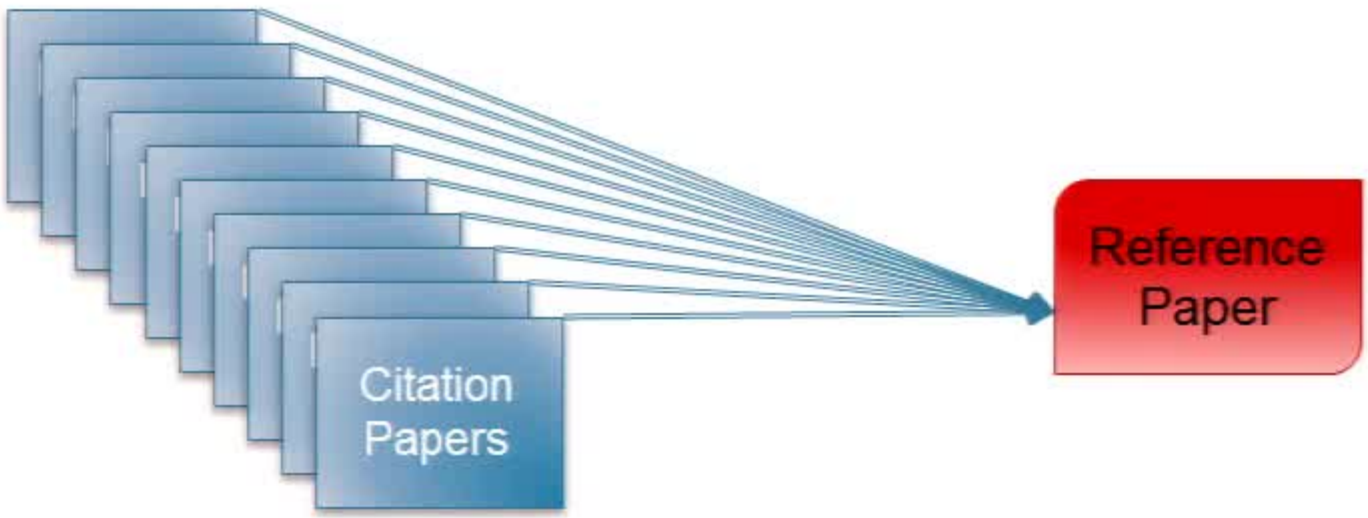
Multilingual Summary Results (100 words)

Selection Method	ROUGE-2	Confidence Interval
Pivoted *QR*	0.156	(0.143,0.169)
NR Factorization	0.179	(0.166,0.192)

Multilingual Summary Results (100 words)

Selection Method	ROUGE-2	Confidence Interval
Pivoted *QR*	0.156	(0.143,0.169)
NR Factorization	0.179	(0.166,0.192)

From News to Science: Summarize a Multiply Cited Paper



Text Analysis Conference (TAC) 2014 Pilot Task:
Biomedical Summarization
20 papers, each with 10 papers citing it.

Vahed Qazvinian and Dragomir R. Radev. COLING 2008 used citing sentences (citances)

At TAC 2014 human were allowed to use citances and reference paper.

Data Processing and Segmentation

- Reference paper is processed,
  - Individual sections of the paper are isolated and extracted, and the document is then sentence split.
  - Sections from reference paper keep:
    - Abstract
    - Results
    - Other parts excluding bibliography.
- For each citation paper one or more citing sentences (*citances*) are extracted.

Term Weighting to Covering

- Resulting term weights, computed by NNMF or LM are correlated with the probability that a human will include the term in a summary.
- *Then, OCCAMS,* an Optimal Combinatorial Covering Algorithm for Multi-document Summarization is then used to select a subset of the sentences, with at most 250 words, with near optimal covering.

Sashka T. Davis, John M. Conroy, and Judith D. Schlesinger. *ICDM Workshops 2012*.

John M. Conroy, Sashka T. Davis, Jeff Kubina, Yi-Kai Liu, Dianne P. O'Leary, and Judith D. Schlesinger. ACL 2013. In *MultiLing Workshop*

OCCAMS_V5 Sentence Selection

- An optimal combinatorial covering algorithm for multi-document summarization.

Algorithm **OCCAMS_V** (T, $\mathcal{D}, \mathcal{W}, c, L$)
1. $K_1 =$ Greedy_BMC$(T, \mathcal{D}, \mathcal{W}, c, L)$ 2. $K_2 = S_{max} \cup$ Greedy_BMC$(T', \mathcal{D}', \mathcal{W}, c', L'))$, where $S_{max} = \text{argmax}_{\{S_i \in \mathcal{D}\}} \left\{ \sum_{t_j \in S_i} w(t_j) \right\}$ and $T', \mathcal{D}', \mathcal{W}, c', L'$ represent quantities updated by deleting sentence S_{max} from the collection. 3. $K_3 =$ KS(Greedy_BMC$(T, \mathcal{D}, \mathcal{W}, c, 5L), L)$; 4. $K_4 =$ KS(K_4', L), where $K_4' = S_{max} \cup$ Greedy_BMC$(T', \mathcal{D}', \mathcal{W}, \mathcal{C}', 5L'))$; 5. $K = \text{argmax}_{k=1,2,3,4} \left\{ \sum_{T(K_i)} w(t_i) \right\}$ where $T(K_i)$ is the set of terms covered by K_i.

Results for NNMF Weights

System	R1	R2	R3	R4
TF	0.511	0.166	0.065	0.030
NNMF_2	0.509	0.172	0.073	**0.036**
NNMF_4	0.504	0.171	**0.074**	0.036
NNMF_35	**0.518**	**0.176**	0.070	0.033
Avg Human	**0.528**	**0.179**	**0.075**	**0.036**
Best Human	0.572	0.219	0.110	0.071

Table 1: Vector Space Model based on TF and NNMF

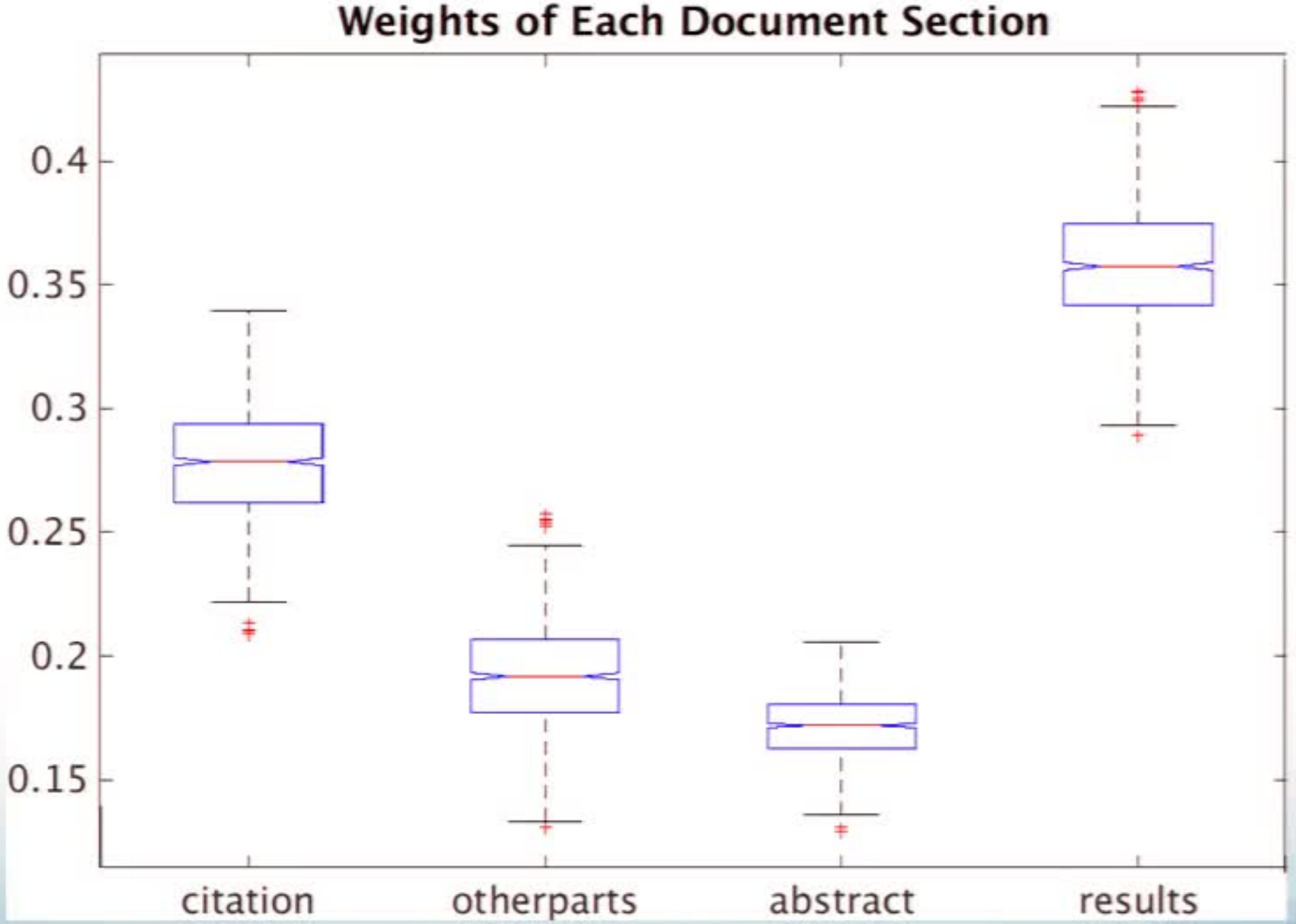
Rankel et al, ACL 2013, R3 and R4 give more discriminant power between systems than R1 or R2.

Results for Language Models

System	R1	R2	R3	R4
LM_1	0.511	0.169	0.067	0.031
LM_4^{equal}	0.559	0.210	0.095	0.052
LM_4^{opt}	0.562	0.216	0.100	0.055
Avg Human	0.528	0.179	0.075	0.036
Best Human	0.572	0.219	0.110	0.071

Table 2: ROUGE Results for Three Language Models and a Comparison to Human Performance

LM Stability of Weights



JSD optimization using 1000 random 10-subsets of the 20 documents sets.