

Bayesian Query-Focused Summarization

Hal Daumé III and Daniel Marcu

Information Sciences Institute
University of Southern California

me@hal3.name, marcu@isi.edu

Three Perspectives

Model for query focused sentence extraction

Well-founded query expansion in LM for IR

Application of LDA-style topic model

Task

Given document, produce summary

important information retained
redundant information removed

What is important to me?

\implies User provides queries

Annotating data is expensive

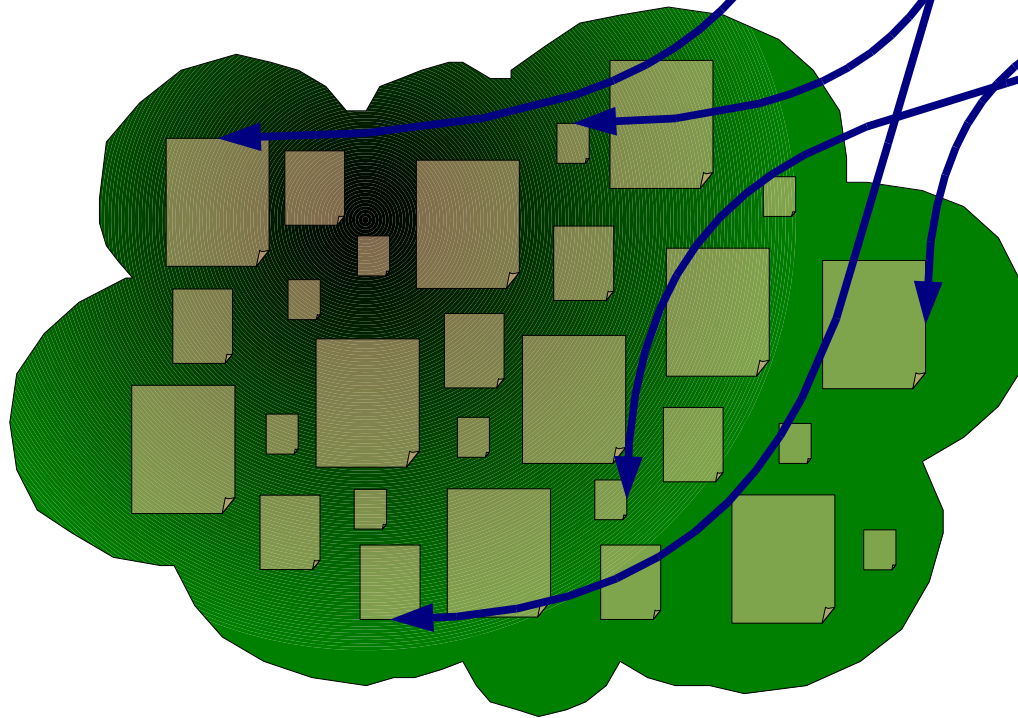
Unsupervised approach (almost)

(caveat: sentence extraction ahead!)

Suppose we have...

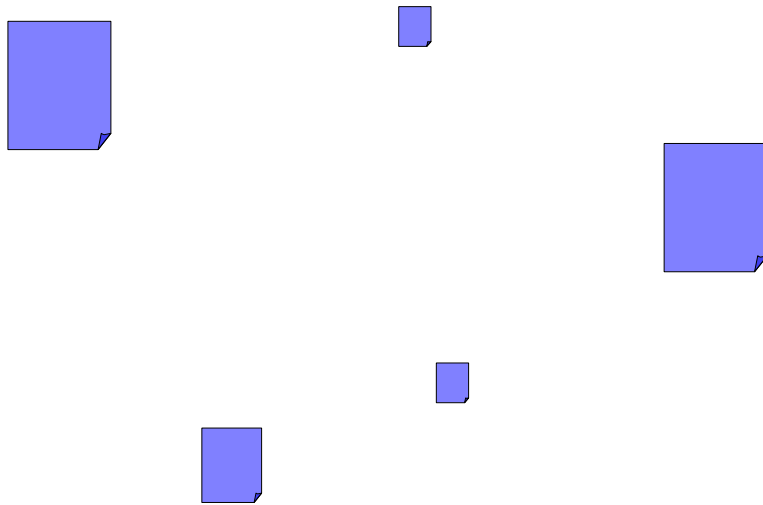
(B) Query “Why is argmax search hard?”

(C) Relevance Judgments



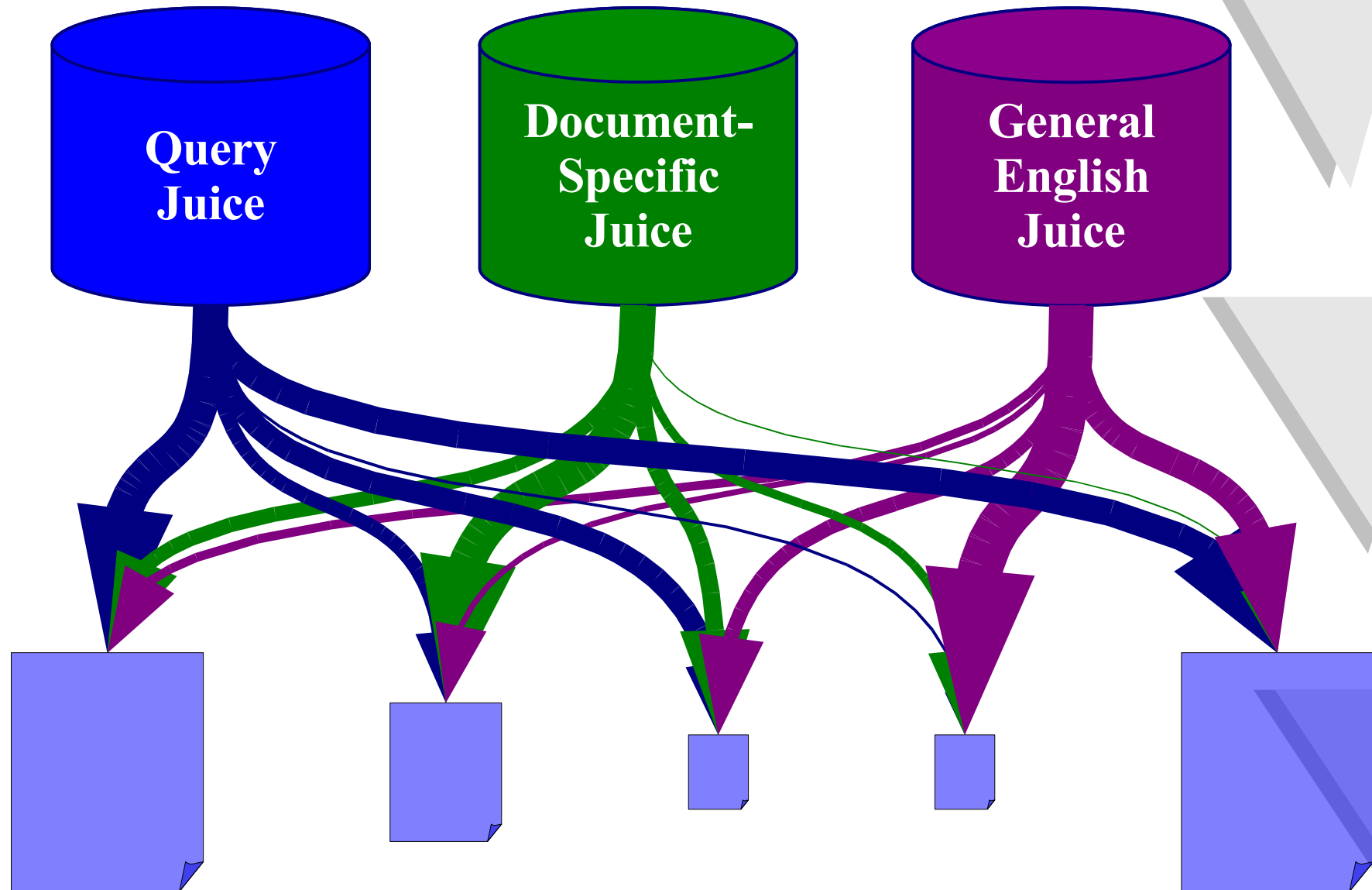
(A) Large Document Collection

Idea



What do relevant documents have that makes them relevant?

A Statistical Model



An Example

Iraq's National Assembly approved a list of Cabinet members for a transitional government Thursday, three months after national elections.

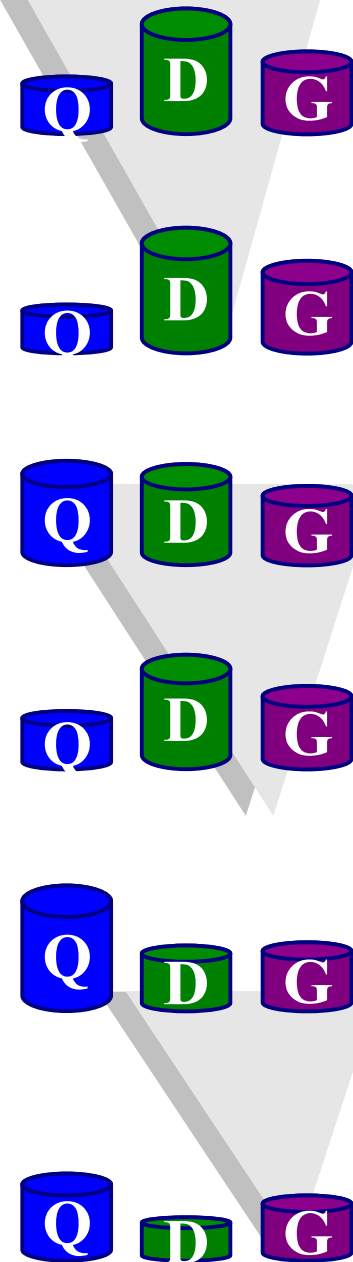
Three ministries – Defense, Oil and Electricity – were filled with temporary appointments because of a last minute failure to reach a compromise.

Prime minister Ibrahim al-Jaafari assumed his post with the creation of his government

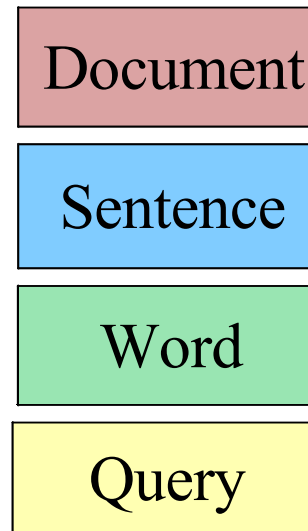
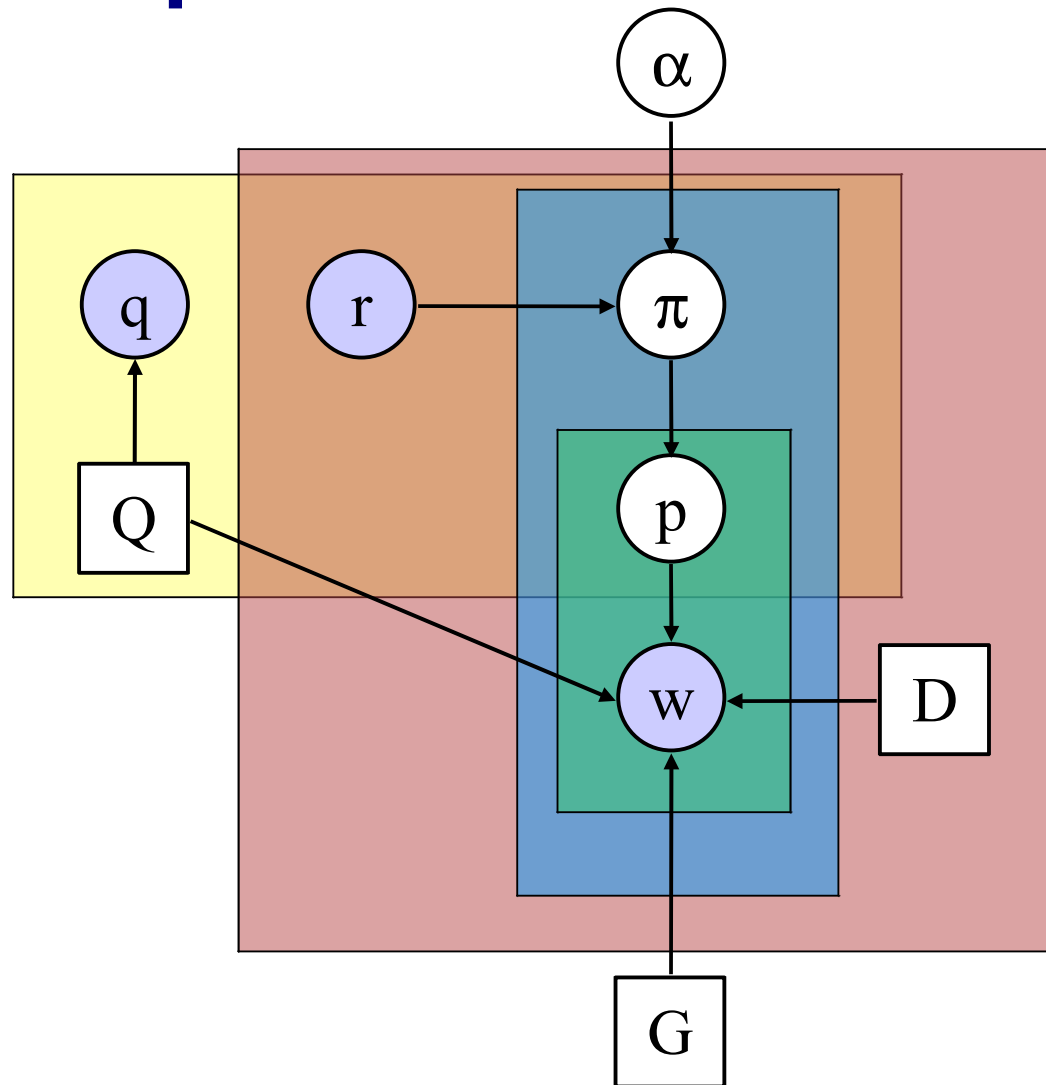
The approval of the Cabinet represents the end of a major political impasse in the country.

On Wednesday, al-Jaafari told a news conference that he had submitted his proposal Cabinet to President Jalal Talabani, who had to approve the names before the transitional National Assembly voted on them.

Al-Jaafari's announcement came a short time after gunmen shot and killed an assembly member on her doorstep in Baghdad...



Graphical Model



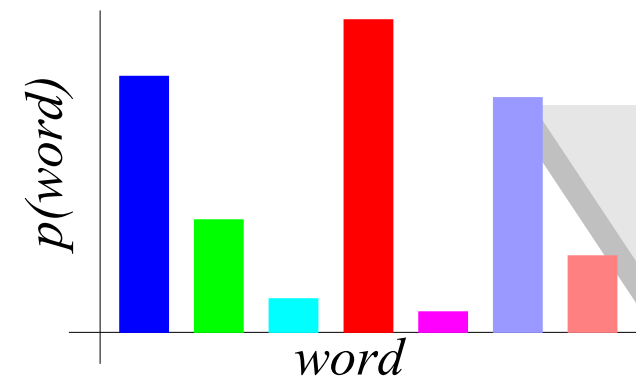
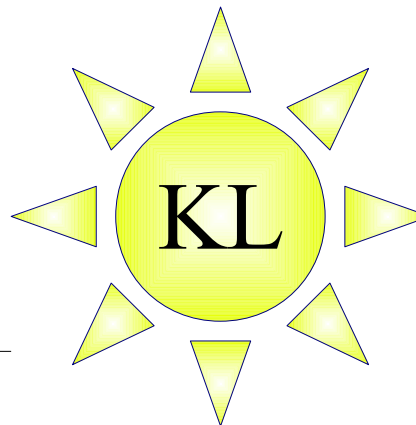
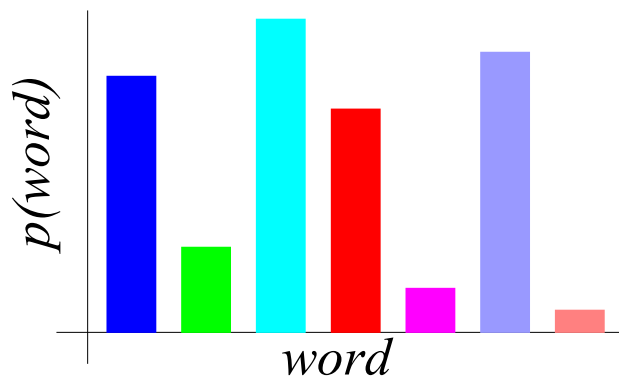
Why does this implement our idea?

$$p(q, r, w) = \prod_j Mult(q_j | Q) \prod_k \prod_m \int_{\Delta} d\pi_{km} Dir(\pi_{km} | \alpha) \prod_n \prod_V \sum_{z_{kmn}} Mult(z_{kmn} | \pi_{km}) Mult(w_{kmn} | V)^{1[V=z_{kmn}]}$$

The LM for IR Perspective...

D O C
 U M E
 N T

Query

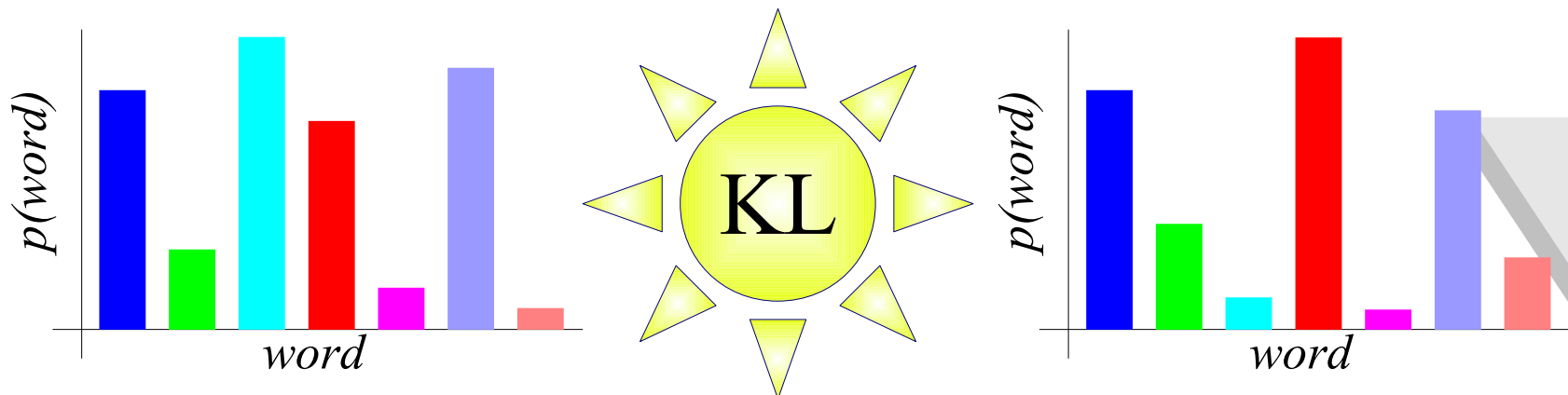


LM for IR in Summaries

Sentence

Query

Sparse \implies Unreliable estimate



Why does it work?

Q: IBM stocks

S1: The price of IBM shares went up 5%

...

S8: Shares in the company are set to top out soon

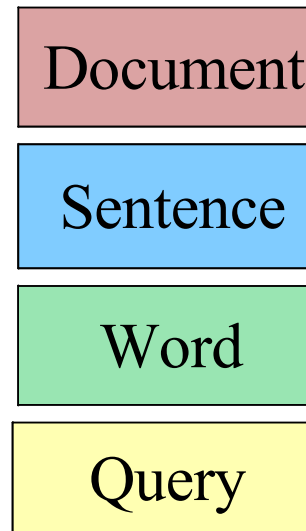
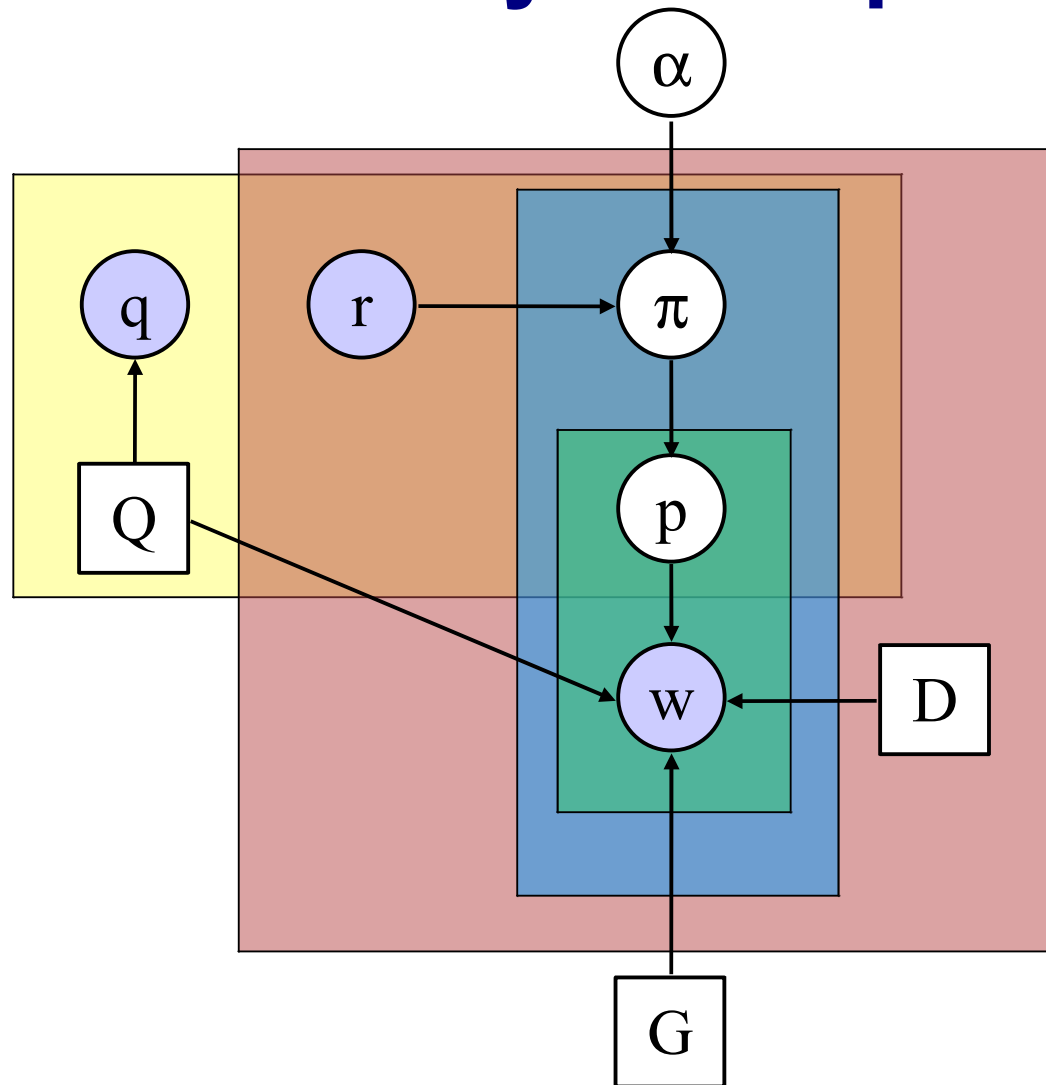
Q: IBM stocks + shares, up, ...

S1: The price of IBM shares went up 5%

...

S8: Shares in the company are set to top out soon

The LDA-style Perspective...



#Q + #D + 1 topics

Constrained topic-to-word mapping

Data

- From the NIST-sponsored Text REtrieval Conference
 - Collection of queries, documents and relevance judgments
 - Queries come in three parts:
 - Title (3-4 words)
 - Description (1 sentence)
 - Narrative (1 short paragraph)
 - We use 350 queries in our experiments
 - All relevant documents:
 - 43 thousand documents, 2.1 million sentences, 65.8 million words
 - Computation time is about 2.5 hours
- Evaluation data:
 - Annotators selected up to 4 sentences for an extract
 - Given title, description and narrative
 - 166 total extracts (reasonably high agreement)
 - Can compute mean average precision, mean reciprocal rank and precision at 2

Baseline Systems

- Strawman models:
 - Rank sentences randomly
 - Rank sentences according to sentence position
 - Compute simple word overlap between query and sentence (Jaccard)
 - Compute simple distance between query and sentence (cosine)
- The real competition
 - State of the art passage retrieval model based on language modeling and KL divergences
 - KL model + blind relevance feedback

Evaluation Results on All Fields

	<u>MAP</u>	<u>MRR</u>	<u>P@2</u>
Random	19.9	37.3	16.6
Position	24.8	41.6	19.9
Jaccard	17.9	29.3	16.7
Cosine	29.6	50.3	23.7
KL	36.6	64.1	27.6
KL+R	36.3	62.9	28.2
BayeSum	44.1	70.8	33.6

Evaluation Results by Field

		MAP	MRR	P@2
Position		24.8	41.6	19.9
Title	KL+R	31.9	53.8	26.1
	BayeSum	41.1	65.7	31.6
+Desc	KL+R	32.6	55.0	26.2
	BayeSum	40.9	66.9	31.0
+Sum	KL+R	34.2	58.5	27.0
	BayeSum	42.0	67.8	31.8
+Nar	KL+R	35.1	60.1	27.3
	BayeSum	42.3	67.5	32.7
+Conc	KL+R	36.3	62.9	28.2
	BayeSum	44.1	70.8	33.6
NO QUERY		39.4	64.7	30.4

Conclusions

Statistical model for query-focused summarization

Inference possible using most any technique

Highly insensitive to size of query

Utilizes relevance judgments in a well-founded way (but robust)

Interesting extensions possible