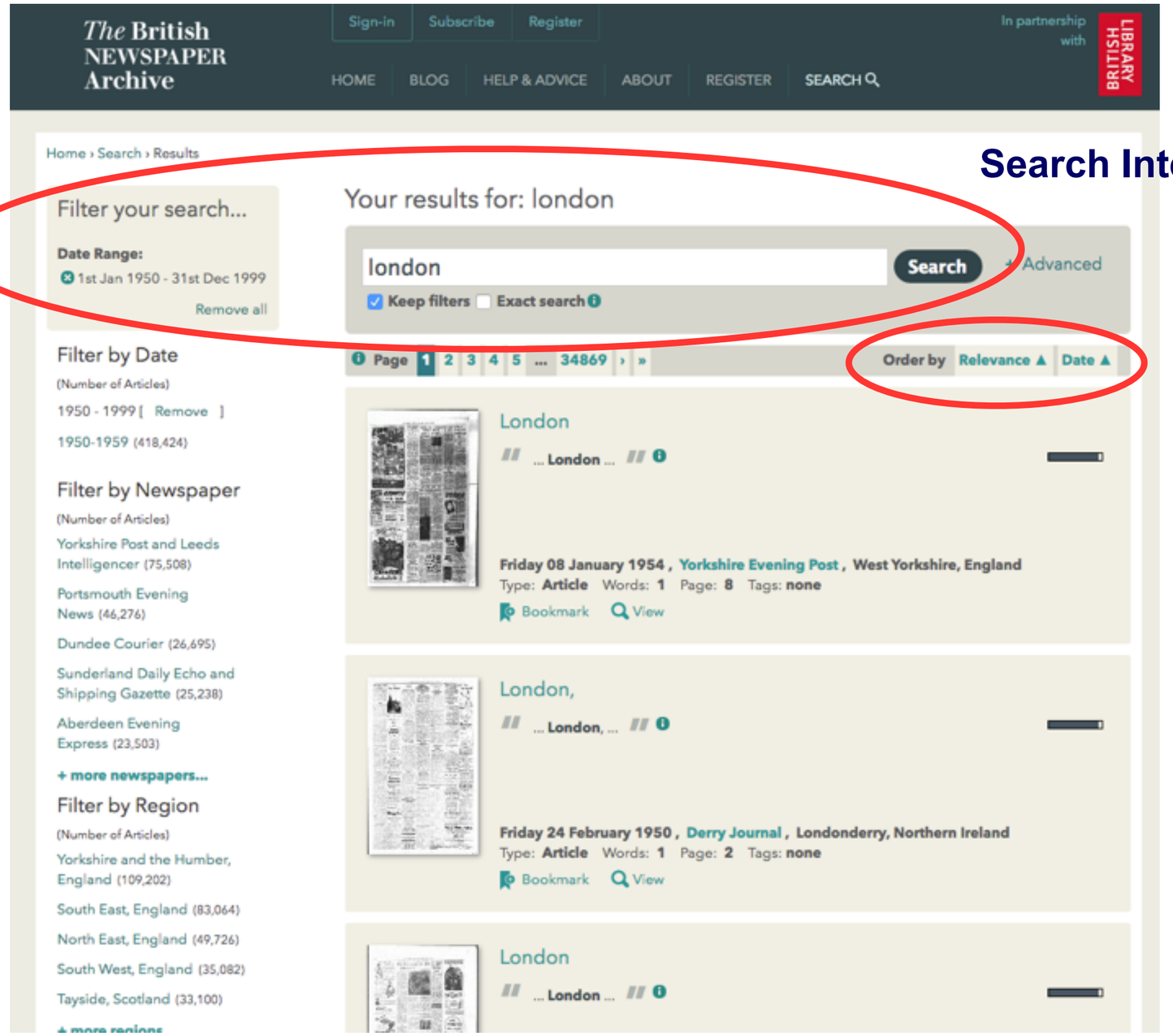
History by Diversity

Jaspreet Singh, Avishek Anand & Wolfgang Nejdl
(singh,anand,nejdl @l3s.de)

Exploring News Archives

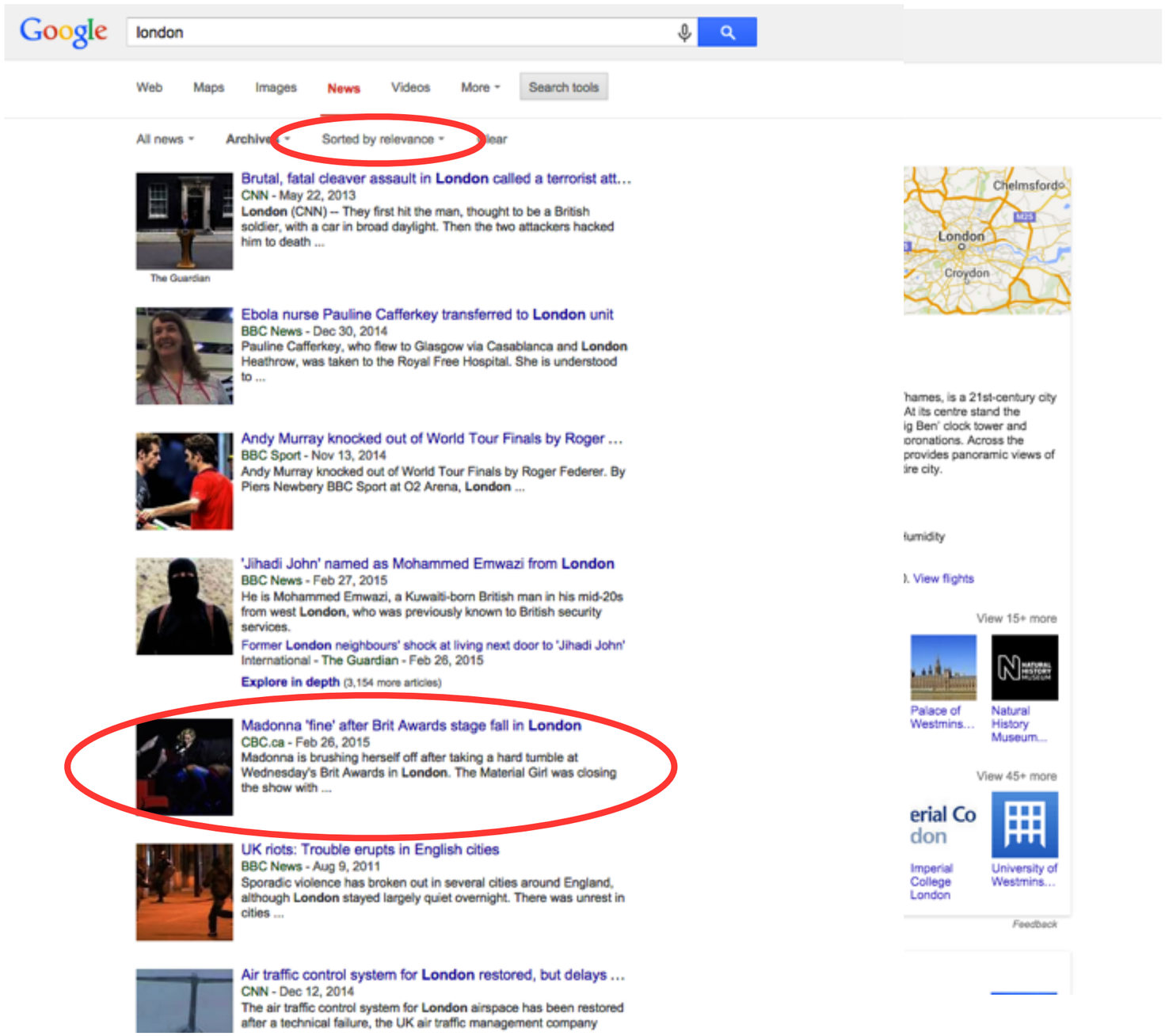


Search Intent

Ranking

Relevant to?

Why can’t it be like Google?

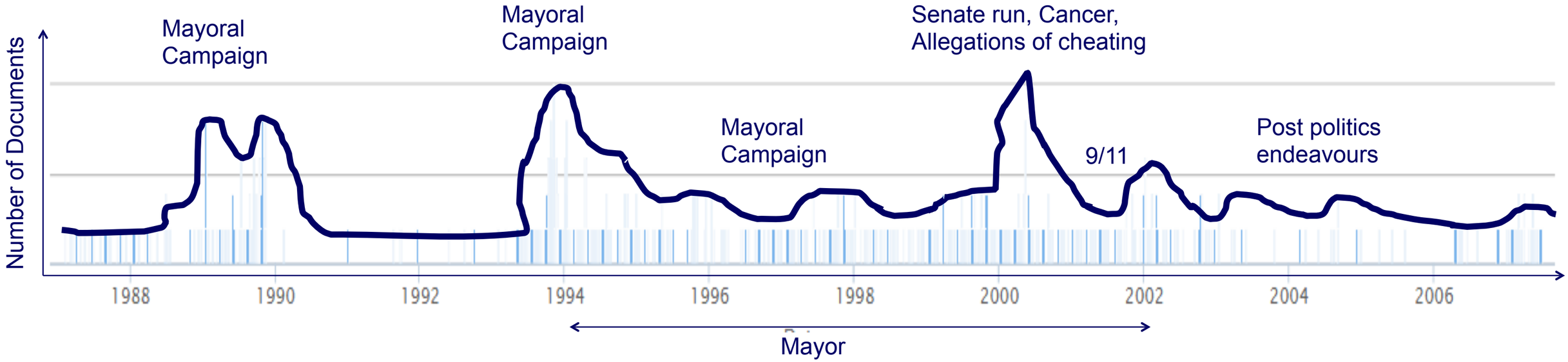


- **Precise**
- **Diverse**
- **Driven by popularity**

What is the intent of a user when using google?

Historical Search on News Archives

I want to know the history of ..... <u>Rudolph Giuliani</u>



- Newspaper articles encode history as it happens.
- Aspects are diverse across time.
- Time windows can be diverse in aspects.

History by Diversity

- Historical Search task: I want documents covering the most important aspects when they were important for a given topic.
- I want documents from the **most important aspects and from the most important time windows**. (New Retrieval Task)

- History by Diversity - Extending the standard diversity problem to include time

$$P(S|q) = \sum_{c} P(c|q)(1 - \prod_{d \in S}(1 - V(d|q, c, \ ))$$

How do you evaluate historical search?

- To measure coverage of important time windows and aspects we introduce a new information space:
- **Aspect-Time space** - encodes which aspects are relevant and when.

SBR@k = $$\frac{\left| \bigcup_{d_p \in \mathcal{R}_q^k} \{ (a_i, \delta_j) \mid a_i \in A(d_p) \wedge \Lambda(p) = \delta_j \} \right|}{|\mathcal{AT}_q|}.$$

- Adapt standard diversity metrics like intent aware precision, subtopic recall, NDCG, etc to function on this space.

- A document is relevant only if it is relevant to an aspect and is also published in its corresponding time period.

Test Collection for Historical Search

- TREC datasets are short time spans; the topics and subtopics are not suited for historical search.

- We created our own test collection using the New York Times 20 year annotated dataset.

- Manually created topics and subtopics using relevant wikipedia history sections.

- Expert binary relevance judgements for 30 topics.

```
<topic>
 <query>rudolph giuliani</query>
 <desc>I want to know the history of Rudolph Giuliani</desc>
 <subtopics>
  <subtopic>
   <desc>Mayoral campaigns</desc>
   <time>[{01.01.1989 – 31.12.1989}, {01.01.1993 – 31.12.1993}, {01.01.97 – 31.12.1997}]</time>
  </subtopic>
  <subtopic>
   <desc>Senate race</desc>
   <time>[{01.01.2000 – 31.12.2000}]</time>
  </subtopic>
  <subtopic>
   <desc>Efforts after 9/11</desc>
   <time>[{11.09.2001 – 01.04.2002}]</time>
  </subtopic>
   .
   .
   .
   .
 </subtopics>
</topic>
```

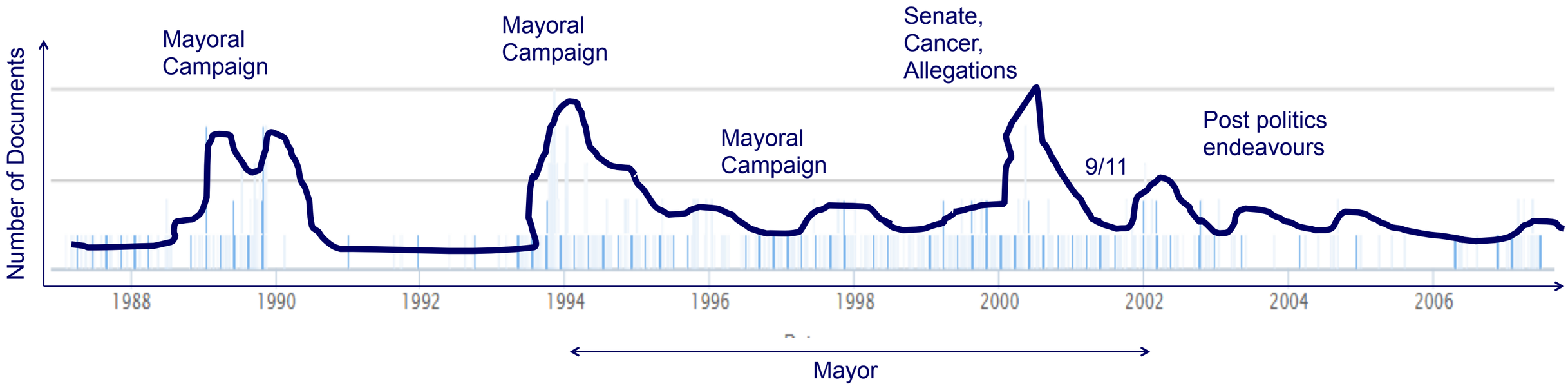
Why not use standard diversification algorithms?

- Diversify just aspects: no guarantee we will get temporal diversity.
- Diversify just time: no guarantee we will get aspect diversity.

- **Aspect Diversity** - Aspect utility is updated using a discounting function.

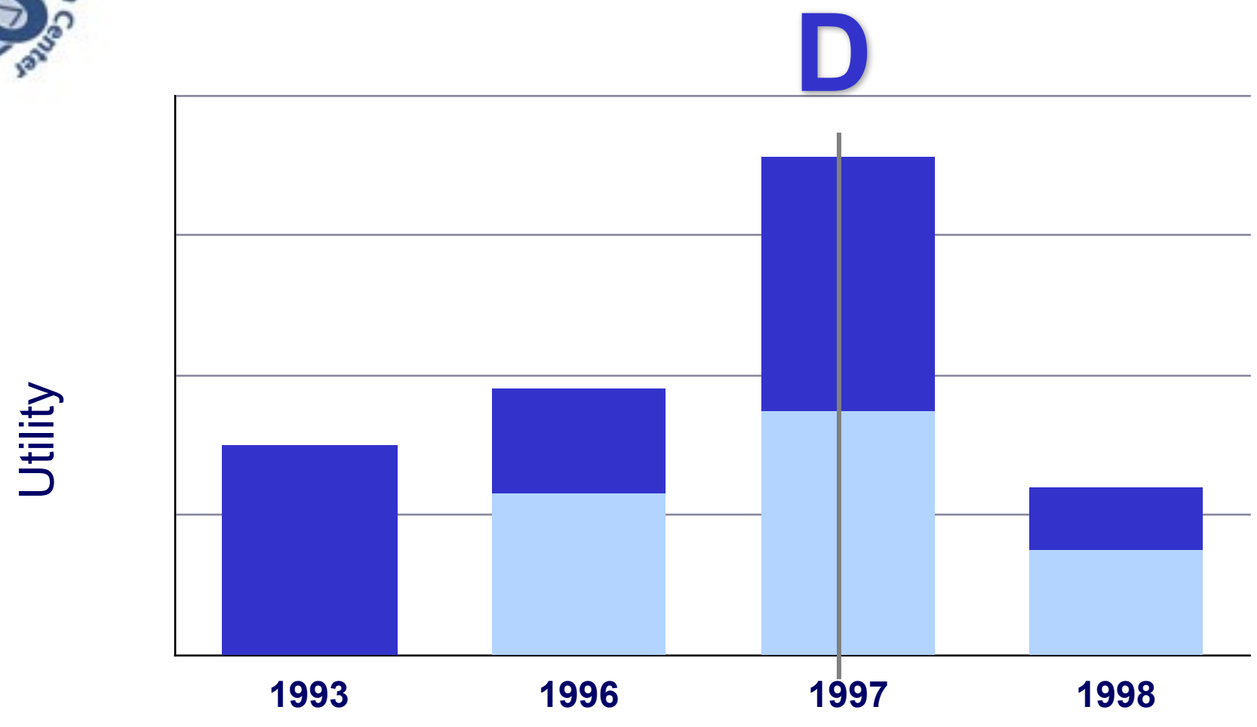
- **Temporal Diversity** - considers aspects of a topic as time windows.
- Discounting time using exponential decay

The HistDiv Approach



- **Aspects are temporal in nature**
- **Time windows are diverse themselves**

- **Compute utility and discount accordingly**



Mayoral Campaign

Mayorality

T = 2000

Only Time

Mayoral Campaign —

Mayorality —

T = 2000 ↓

Only Aspect

Mayoral Campaign ↓

Mayorality —

T = 2000 —

HistDiv

Mayoral Campaign ↓

Mayorality —

T = 2000 ↓

The HistDiv Algorithm

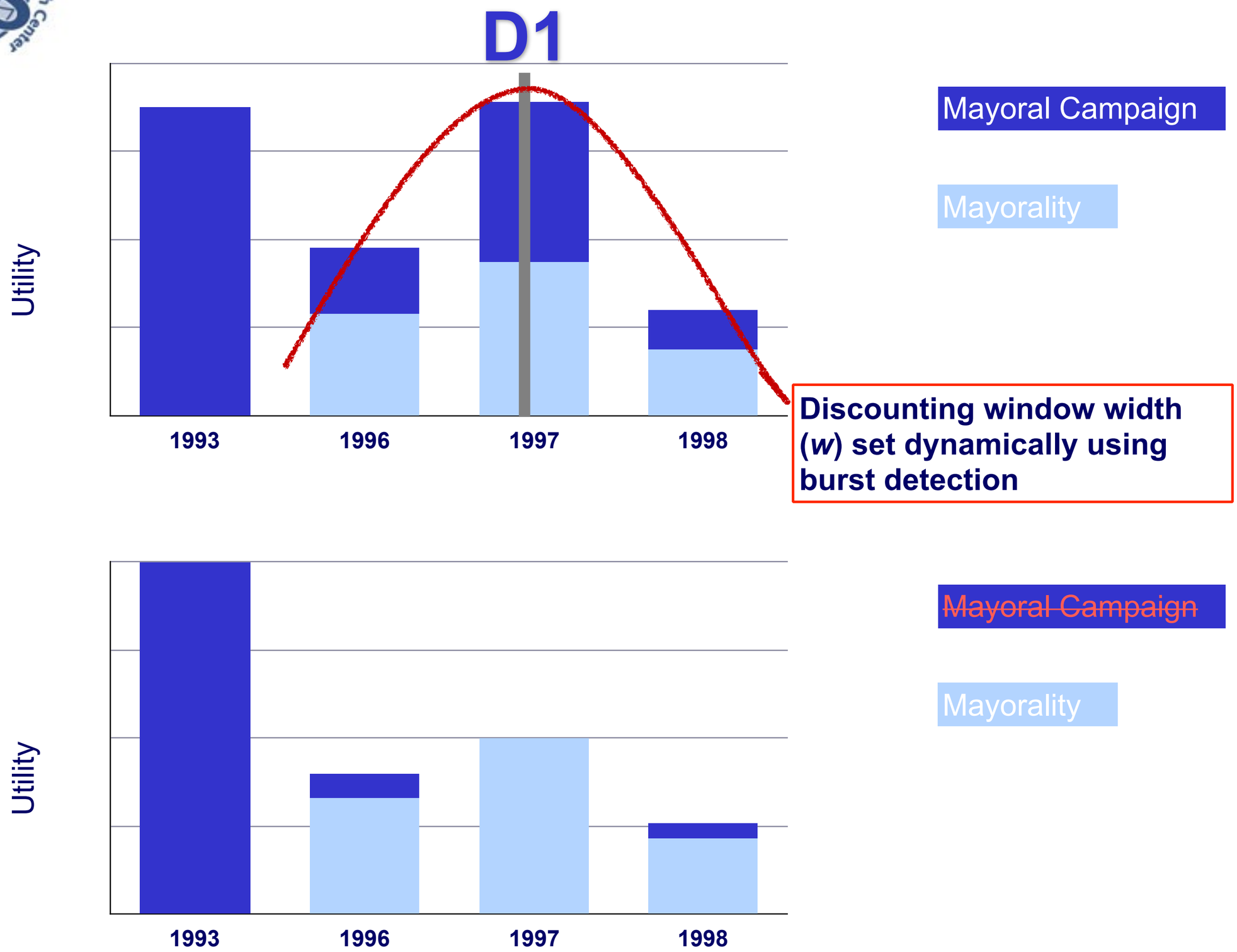
- Extend the multi-dimension diversification algorithm
- 2 dimensions: Time and Aspects

$$g(d|q,S) \leftarrow \alpha.V(d|q) + (1-\alpha).(\beta.\sum_{c}^{A(d)} U_{aspect} + (1-\beta).U_{time})$$

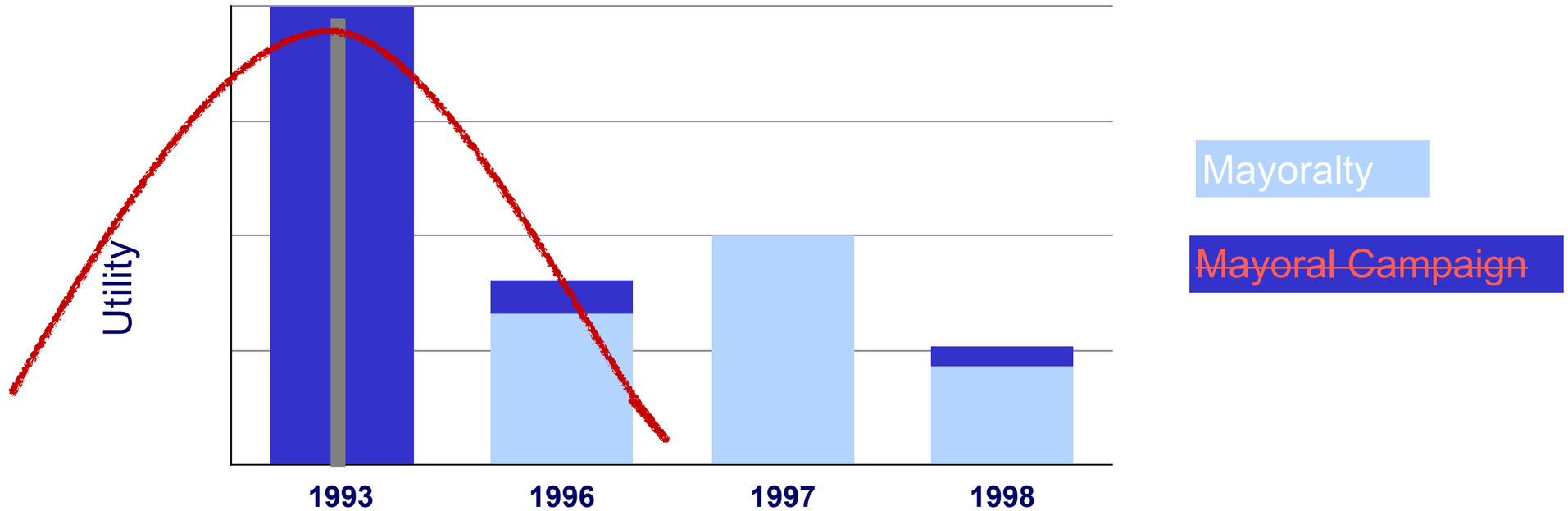
Time Decay based discounting

Coverage based discounting

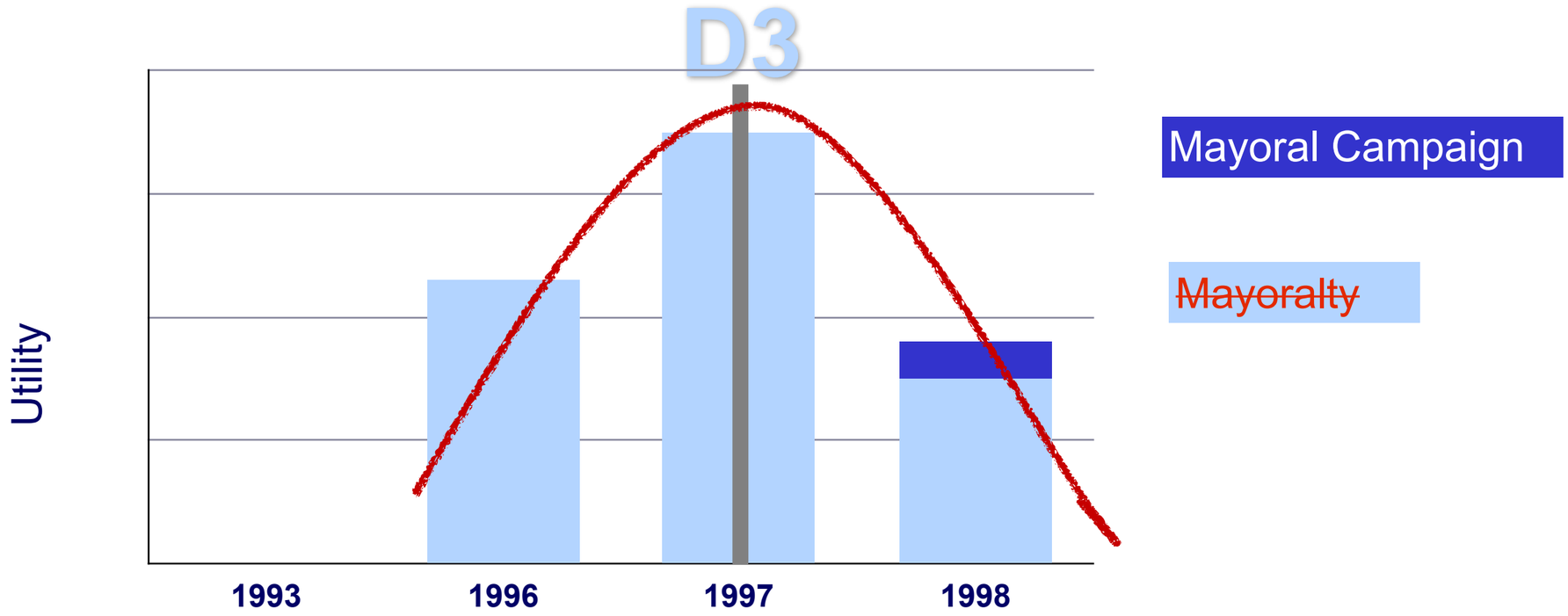
- Dimension are **interdependent**


D1
Utility
1993
1996
1997
1998
Mayoral Campaign
Mayorality
Discounting window width (*w*) set dynamically using burst detection
Utility
1993
1996
1997
1998
~~Mayoral Campaign~~
Mayorality

D2


Utility
1993
1996
1997
1998
Mayoralty
~~Mayoral Campaign~~

D3


Utility
1993
1996
1997
1998
Mayoral Campaign
~~Mayoralty~~

Experiments

- New York Times Test Collection
- Metrics: Time Aware NDCG, ERR, Subtopic Recall, Precision, MAP & TIA-SBR
- Window size: year & month
- Aspects mined from AIDA & wikiminer
- Tuned for best performance in subtopic recall.

Competitors

- Competitors - Time diversification, Aspect Diversification, Adapted aspect diversification & Multi Dimension Diversification

- We strengthen the following competitors by linearizing aspects and time

  - IA-SELECT
  - PM2

  Other competitors: Non temporal IA-SELECT & PM2, MDIV, OnlyTime

  Baseline: Language Model with dirichlet smoothing

Results

	k=10			k=15			k=20		
	A	T	AT $(W/L\%)$	A	T	AT $(W/L\%)$	A	T	AT $(W/L\%)$
LM	0.706	0.060	0.428	0.752	0.085	0.491	0.780	0.091	0.518
IA-SELECT$^{\circ}$	0.722	0.039	0.442 (23/23)	0.766	0.047	0.491 (20/26)	0.841	0.055	0.516 (20/23)
PM2*	0.707	0.069	0.429 (16/20)	0.794	0.082	0.471 (10/23)	0.817	0.097	0.509 (16/26)
TIA-SELECT$^{\bullet}$	0.614	0.039	0.380(23/36)	0.717	0.047	0.433 (20/43)	0.770	0.055	0.470 (20/26)
T-PM2$'$	0.551	**0.088**	0.308 (13/50)	0.680	0.106	0.408(20/43)	0.761	0.128	0.453 (16/33)
E-IA-SELECT‡	0.700	0.062	0.435 (23/23)	0.776	0.084	0.501 (23/23)	0.837	0.095	0.524 (23/20)
E-PM2†	0.692	0.061	0.422 (6/16)	0.766	0.083	0.469 (6/26)	0.816	0.098	0.495 (10/26)
EQT	0.714	0.076	0.440 (16/13)	0.766	0.097	0.503 (13/6)	0.802	0.117	0.542 (20/6)
MDIV$^{\blacktriangle}$	0.720	0.060	0.460 (33/33)	0.764	0.079	0.515 (23/16)	0.823	0.096	0.552 (29/3)
ONLYTIME$^{\diamond}$	0.729	0.068	0.426 (20/26)	0.807	0.092	0.497 (26/26)	0.826	0.115	0.534 (26/20)
HISTDIV	0.761$^{\diamond}$	0.07	0.497$^{\blacktriangle}$ (40/13)	0.814	0.085	0.542$^{\blacktriangle}$ (36/26)	**0.864**‡	0.101	0.583$^{\blacktriangle}$(43/13)
HISTDIV-BURST	**0.777**$^{\diamond}$	0.087	**0.509**$^{\blacktriangle}$ (33/6)	**0.830**$^{\diamond}$	**0.113**$'$	**0.560**$^{\blacktriangle}$ (46/20)	0.860‡	**0.132**	**0.601**$^{\blacktriangle}$ (43/16)
HISTDIV-NER	0.741	0.110	0.467				0.862	0.104	0.588
HISTDIV-BURST-NER	0.761	0.137	0.483				0.840	0.140	0.561
NYT	0.473	0.046	0.288	0.552	0.057	0.329	0.578	0.062	0.346
GOOGLE	0.564	0.068	0.312	0.621	0.077	0.353	0.663	0.085	0.402

TIA-SBR (Win/Loss)

	IAP		SBR		NDCG		IA-ERR		MAP	
	M	Y	M	Y	M	Y	M	Y	M	Y
LM	0.099	0.099	0.428	0.428	0.402	0.402	0.201	0.201	0.228	0.228
IA-SELECT$^{\circ}$	0.101	0.101	0.442	0.442	0.415	0.415	0.180	0.180	0.215	0.215
PM2*	0.100	0.100	0.429	0.429	0.388	0.388	0.213	0.213	0.241	0.241
TIA-SELECT$^{\bullet}$	**0.120**$^{\blacktriangle}$	**0.113**‡	0.380	0.361	**0.497**‡	**0.468**$^{\circ}$	0.195	0.179	0.242	0.232
T-PM2$'$	0.064	0.091	0.308	0.410	0.232	0.368	0.123	0.176	0.152	0.167
E-IA-SELECT‡	0.106	0.102	0.435	0.430	0.478	0.412	0.183	0.177	0.219	0.214
E-PM2†	0.103	0.099	0.422	0.417	0.419	0.379	0.217	0.204	0.227	0.239
EQT	0.096	0.078	0.441	0.426	0.360	0.331	0.203	0.200	0.229	0.213
MDIV$^{\blacktriangle}$	0.109	0.096	0.460	0.428	0.389	0.370	0.204	0.203	0.236	0.236
ONLYTIME$^{\diamond}$	0.089	0.076	0.426	0.415	0.354	0.297	0.196	0.189	0.236	0.220
HISTDIV	0.096	0.087	0.497$^{\blacktriangle}$	0.459°	0.383	0.339	0.229*	0.208	**0.255**$^{\bullet}$	0.231
HISTDIV-BURST	0.096	0.096	**0.509**$^{\blacktriangle}$	**0.509**$^{\circ}$	0.375	0.375	**0.231***	**0.231***	0.244	**0.244**
HISTDIV-NER	0.097	-	0.464	-	0.391	-	0.213	-	0.245	-
HISTDIV-BURST-NER	0.091	-	0.483	-	0.358	-	0.210	-	0.225	-
NYT	0.055	-	0.288	-	0.206	-	0.126	-	0.154	-
GOOGLE	0.059	-	0.312	-	0.216	-	0.147	-	0.225	-

Conclusion

- ✓ Historical Query Intents, test collection to evaluate retrieval models for HQIs and a new metric TIA-SBR.
- ✓ HistDiv Algorithm - special semantics to discount time and aspects
- ✓ Outperform competitors in most measures.
- ✓ Aspects and time are interlocked
- ✓ Robust - temporal references alone are also effective as well as simple NER
- ✓ Recall at the cost of precision
- ✓ Good starting point for further exploration of a news archive

Demo

http://pharos.l3s.uni-hannover.de:7080/ArchiveSearch/starterkit/