



SECTION 4 · 25 min

Foundations: Pre-LLM Temporal IR

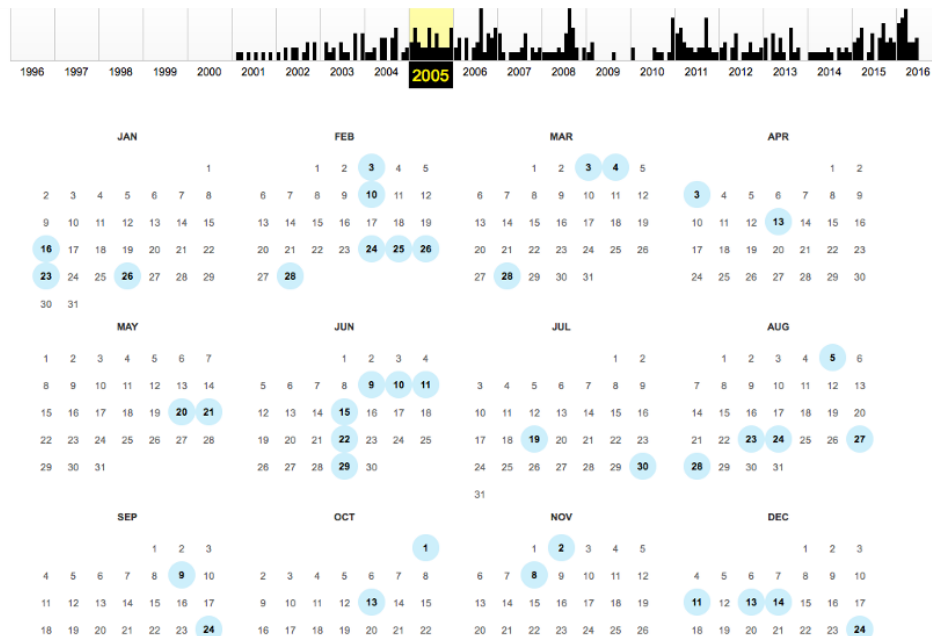
Rule-based systems and statistical models that laid the groundwork — and the principles that still matter.

 *Presenter: Avishek Anand*

0

3

Internet Archive: Wayback Machine



Only lookups allowed, historical queries not supported

Value of Temporal Collections

Temporal Collections have great value

- Historical information needs
- Captures evolution
- **Example:** *The history NATO interventions in the last 100 years*

Tunis Asks **NATO** Intervention to End French-Algerian Conflict

TUNIS, Feb. 27. (AP)—Tunisian President Habib Bourguiba today called on the NATO powers to intervene to end the conflict between France and the Algerian rebels.

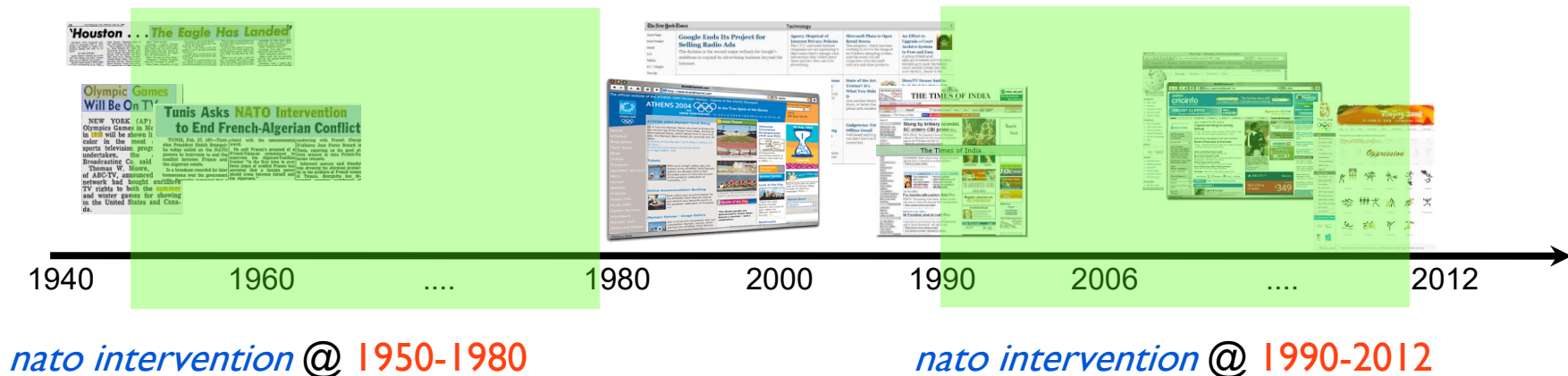
ciated with the noncommunist world.

He said France's proposal of a French-Tunisian commission to supervise the Algerian-Tunisian frontier "is the first time in over three years of conflict France has

conferring with French Charge D'affaires Jean Pierre Benard in Tunis, reporting on his good offices mission to ease French-Tunisian tensions.

Informed sources said Murphy

Time-Travel Search



[Tunis Asks Nato Intervention To End French-algerian...](#)

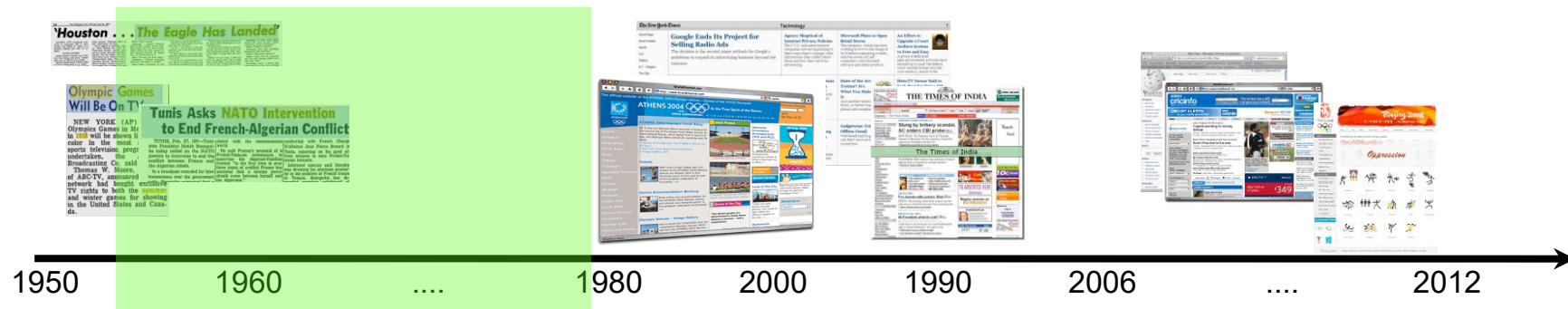
Spokane Daily Chronicle - Feb 27, 1958

Habib Bourguiba today called on the **NATO** powers to **intervene** to end the conflict between France and the Algerian rebels. In a broadcast recorded for later ...

[Nato .Intervention Ashed In French....](#) Reading Eagle

Search workloads interesting and important

Time-Travel Search



nato intervention @ 1950-1980

Tunis Asks Nato Intervention To End French-algerian...

Spokane Daily Chronicle - Feb 27, 1958

Habib Bourguiba today called on the **NATO** powers to **intervene** to end the conflict between France and the Algerian rebels. In a broadcast recorded for later ...

Nato .Intervention Ashed In French.... Reading Eagle

Search as a primitive operation

Wikipedia Revision History

Help: Page history, Revision history

View logs for this page

Browse history

From year (and earlier): From month (and earlier): all Tag filter: Deleted only ☐ Go

For any version listed below, click on its date to view it. For more help, see [Help:Page history](#) and [Help:Edit summary](#).

External tools: [Revision history statistics](#) · [Revision history search](#) · [Edits by user](#) · [Number of watchers](#) · [Page view statistics](#)

(cur) = difference from current version, (prev) = difference from preceding version, m = minor edit, - = section edit, - = automatic edit summary

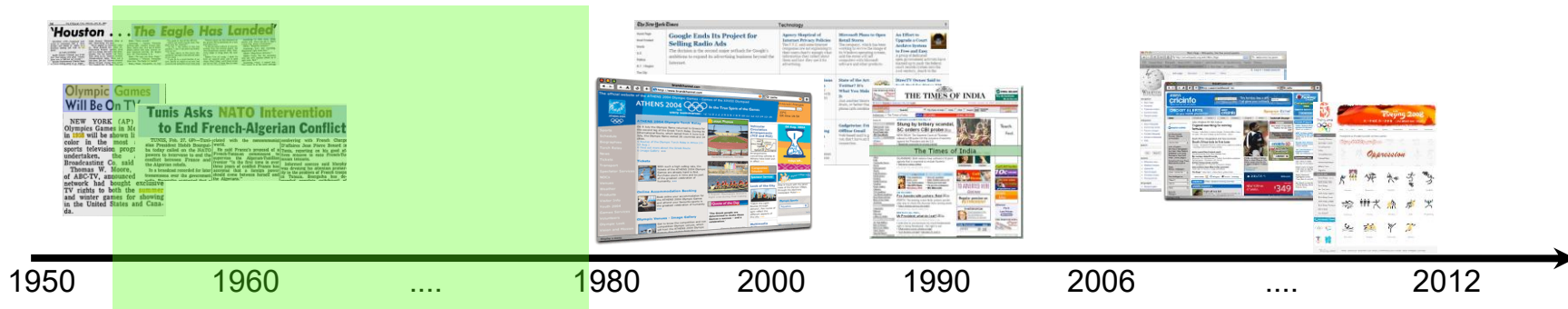
(latest | earliest | View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500))

Compare selected revisions

- (cur | prev) 16:48, 31 October 2012 Whatamidoing (talk | contribs) m . . (16,289 bytes) (+1) . . (Type)
- (cur | prev) 16:47, 31 October 2012 Whatamidoing (talk | contribs) . . (16,288 bytes) (+599) . . (Once again, with feeling)
- (cur | prev) 23:45, 28 October 2012 Whatamidoing (talk | contribs) . . (15,689 bytes) (+133) . . { - See also: *AR*}
- (cur | prev) 20:43, 24 September 2012 Whatamidoing (talk | contribs) . . (15,556 bytes) (+370) . . (Expand lead)
- (cur | prev) 18:19, 11 September 2012 The wub (talk | contribs) . . (15,186 bytes) (-33) . . (rm {{Leave feedback}}, has article feedback tool now instead)
- (cur | prev) 16:36, 25 July 2012 John of Reading (talk | contribs) . . (15,219 bytes) (-11,036) . . (Reverted 1 edit by Ace JerZe (talk); Rv mislabeled draft, if that's what it was. (TW))
- (cur | prev) 16:13, 25 July 2012 Ace JerZe (talk | contribs) . . (26,255 bytes) (+11,036)
- (cur | prev) 20:47, 5 June 2012 Arcandam (talk | contribs) . . (15,219 bytes) (+5) . . (vector)
- (cur | prev) 22:42, 5 May 2012 Sefid par (talk | contribs) . . (15,214 bytes) (+44) . . (Undid revision 490861562 by Sefid par (talk))
- (cur | prev) 22:41, 5 May 2012 Sefid par (talk | contribs) . . (15,170 bytes) (-44) . . (taoult16)
- (cur | prev) 09:33, 10 April 2012 Elphibot (talk | contribs) m . . (15,214 bytes) (+33) . . (2.7.2) (Robot: Adding no help:Revisionhistory)
- (cur | prev) 07:16, 22 March 2012 JAnBot (talk | contribs) m . . (15,182 bytes) (0) . . (2.5.4) (Robot: Modifying sr:Πομπή Ιστοπομπής Ιστομετρία, sr:Υπόμνημα Sayfa)
- (cur | prev) 11:04, 7 March 2012 John of Reading (talk | contribs) . . (15,182 bytes) (+1) . . { - Overview: r.v.v.o old description of the edit summary. The screenshot now has a separate marker for the section edit summary.
- (cur | prev) 22:42, 3 March 2012 PrimeHunter (talk | contribs) . . (15,181 bytes) (+105) . . { - Overview: clarification.
- (cur | prev) 10:29, 3 March 2012 John of Reading (talk | contribs) . . (15,076 bytes) (+246) . . { - Overview: Filled in one of the missing explanations)
- (cur | prev) 09:56, 3 March 2012 John of Reading (talk | contribs) . . (14,830 bytes) (+336) . . { - Overview: New screenshot and some changes to the descriptions. Parts of the descriptions need more work.)
- (cur | prev) 16:48, 2 March 2012 SQ6 (talk | contribs) . . (14,494 bytes) (+10) . . (Outdated flag. Needs a new screenshot with the new green & red parenthetical info. Remove when accomplished. Thanks. See talk page.)

Evolution of knowledge, foundation for many KBs, used in many linking tasks

Temporal Indexing Setup



nato intervention @ 1950-1980

- **Temporal Collection** – Web archive pages, news archives, Wikipedia versions, emails, etc.
- **Temporal Query** – Interested in a time point or interval in the past
- **Results** – Documents/versions satisfying constraints in the query

Temporal Representation of documents

Discrete Time: Integers as timestamps

"now": Always points to the current time

Document d: Sequence of time-stamped versions

Deletion: Results in tombstone version $\perp$

Valid-time interval: $\text{val}(\text{dti})$

State of Document Collection D during $[t_b, t_e]$

$$D^{[t_b, t_e]} = \bigcup_{d \in D} \left\{ d^{t_i} \in d \mid \wedge \begin{array}{l} d^{t_i} \neq \perp \\ [t_b, t_e] \cap \text{val}(d^{t_i}) \neq \emptyset \end{array} \right\}$$

Temporal Indexing

Given a versioned collection of documents with valid time intervals also called Time-travel text queries

Query Example

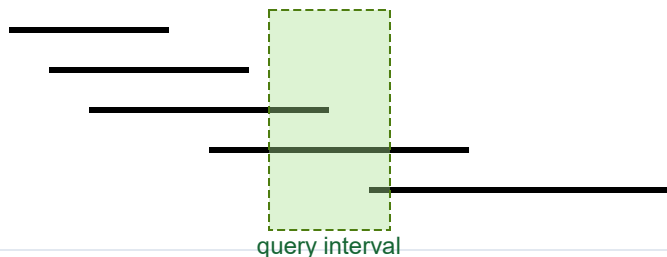
nato intervention @ 2001 - 2004

Boolean Temporal Retrieval: We want to retrieve documents containing terms **“nato”**, **“intervention”** and valid between *2001 - 2004*

How do we efficiently retrieve these documents?

Temporal Indexing

temporal distribution



(d1, [t1, t9))
(d2, [t2, t8))
(d3, [t3, t6))
(d4, [t4, t6))
(d5, [t5, t7))

Data Model: Each document is associated with a valid time interval

Query Model: Queries are associated with a time interval [tb, te]

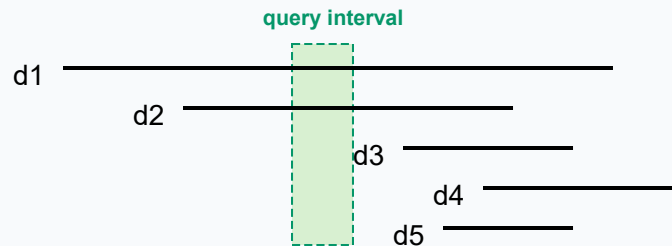
- Point in time queries: when begin time = end time
- Time interval queries

Temporal Indexing

Add temporal information to postings lists in an inverted index

- Each Posting for a term t has docid as well as a time interval
- Filter postings which are outside the time interval

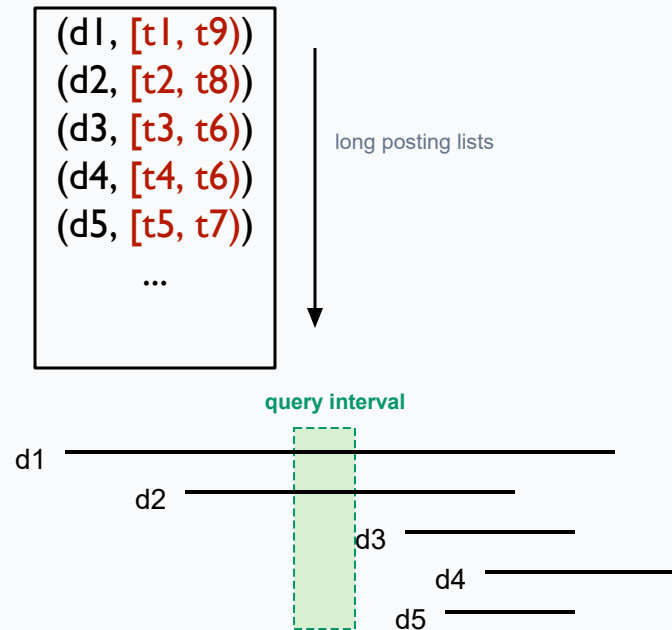
$(d1, [t1, t9])$
 $(d2, [t2, t8])$
 $(d3, [t3, t6])$
 $(d4, [t4, t6])$
 $(d5, [t5, t7])$
...



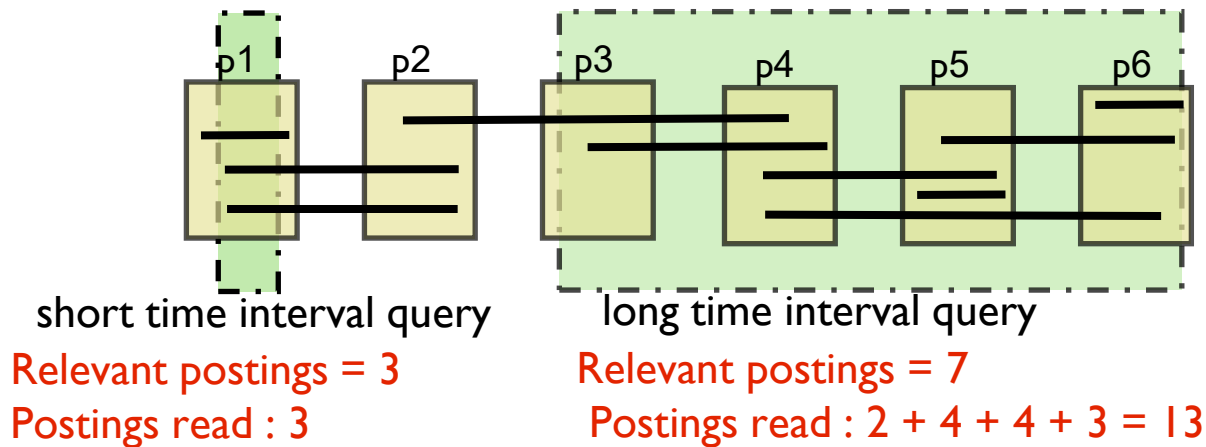
Temporal Indexing

Efficiency Mandate

- Avoid unwanted or wasted access to posting lists
- Access only relevant postings (bounded loss)
- Time points are handled as range queries (sorting by begin time)



Temporal List Partitioning



Short Query: Avoids reading/accessing full postings list – **partitioning good**

Long queries: Is problematic because of redundant reads – **partitioning bad**

Challenges in Temporal Indexing

- Index Partitioning
 - Can we **partition** the indexing for faster temporal querying ?
 - How do we maintain indexes in case of incremental updates ?
- Index Compression
 - How does one exploit redundancy to reduce index size ?
 - **Lossless** vs **Lossy** compression

[Anand et al. CIKM 10; SIGIR 11,12; He SIGIR 11,12]

Temporal Information Needs

Information needs with a **temporal dimension**

*"FIFA World Cup tournaments **of the 1990's**"*
*"Movies that won an Academy Award **in 2007**"*
*"Crusades **of the 12th century**"*
*"London Summer Olympics **2012**"*

Queries containing **temporal expressions** (e.g., **in 1998**)

- Indicate an **underlying temporal information need**
- Account for **1.5% of general web queries**
- More common for **specific domains** (News/Sports) or **user groups** (journalists/historians)

But: Not well-supported by existing retrieval models

Temporal Intents

- Existing retrieval models don't perform well for **information needs** that have a **temporal dimension**

fifa world cup
1990's



In the last decade the FIFA World Cup was won by Germany (in 1990), Brazil (in 1994), and France (in 1998).

- Idea:** Leverage **temporal expressions** (e.g., **in 1998**) contained in documents and queries
- Challenges:** **Meaning** of temporal expressions **uncertain**, not clear how to **seamlessly integrate** them into a retrieval model

Temporal expressions

- Temporal expressions can be categorized as:
 - **Explicit** (e.g., July 19th 2010 or September 1872)
 - **Implicit** (e.g., Christmas 2009 or New Year's Eve 2000)
 - **Relative** (e.g., yesterday or last month)
- Meaning of a temporal expression is often uncertain, e.g.,
- **Document and Query Model:** Distinction between the textual part and the temporal part of documents and queries
 - d_{text} and q_{text} are bags of textual terms
 - d_{time} and q_{time} are bags of temporal expressions

Temporal Expressions Model

We **model** a temporal expression as a **four-tuple**:

$$T = (tb_l, tb_u, te_l, te_u)$$

It records the **earliest/latest begin/end** of any time interval $[b, e]$ that T may refer to.

Example: "in 1998"

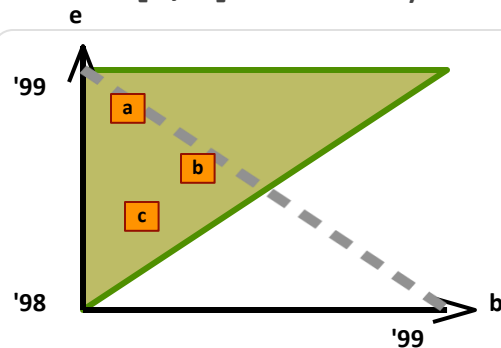
Represented as:

(1998/01/01, 1998/12/31, 1998/01/01, 1998/12/31)

(a) [1998/01/01, 1998/12/31]

(b) [1998/07/12, 1998/07/12]

(c) [1998/02/07, 1998/02/22]



[Berberich et al., ECIR 2010]

Temporal Language Models

1. Query-Likelihood Approach $P(q | d) = P(q_{\text{text}} | d_{\text{text}}) \times P(q_{\text{time}} | d_{\text{time}})$

Assuming that the **textual** and **temporal** query parts are generated *independently*.

2. Independent Generation

$$P(q_{\text{time}} | d_{\text{time}}) = \prod_{Q \in q_{\text{time}}} P(Q | d_{\text{time}})$$

Generation of query temporal expressions as a product of individual probabilities.

3. Two-Step Generation of Q

- (1) Draw a temporal expression **T** at **uniform random**.
- (2) Generate **Q** from **T**.

$$P(Q | d_{\text{time}}) = \frac{1}{|d_{\text{time}}|} \sum_{T \in d_{\text{time}}} P(Q | T)$$

[Berberich et al., ECIR 2010]

Temporal Language Models

Conceptual Intuition: Intuitively, $P(Q|T)$ reflects the probability that the user issuing the query and the author writing the document have the same time interval in mind.

Simplified Definition: The definition of $P(Q|T)$ can be simplified as treating Q and T as sets of time intervals.

Computational Efficiency: $|T|$, $|Q|$, and $|T \cap Q|$ efficiently computable based on the four-tuple representation, i.e., no need to enumerate the huge but finite number of time intervals in Q and T .

$$P(Q|T) = \frac{|T \cap Q|}{|T| \cdot |Q|}$$

Freshness In Rankings

Introduction to Freshness

Freshness of documents could be utilised for improving ranking.

How do we include freshness?

Querying / Hitting Time

The specific time when the query was issued.

Recency & Relevance

- Recently published documents are more relevant than older ones.
- Exponential decay is used to model old documents.



Temporal Diversification

“I am looking for **relevant** documents regarding the most important **aspects** from the **time** period when these aspects were relevant.”

Relevance

The degree to which a document matches the user's information need based on topicality and temporal constraints.

Aspects

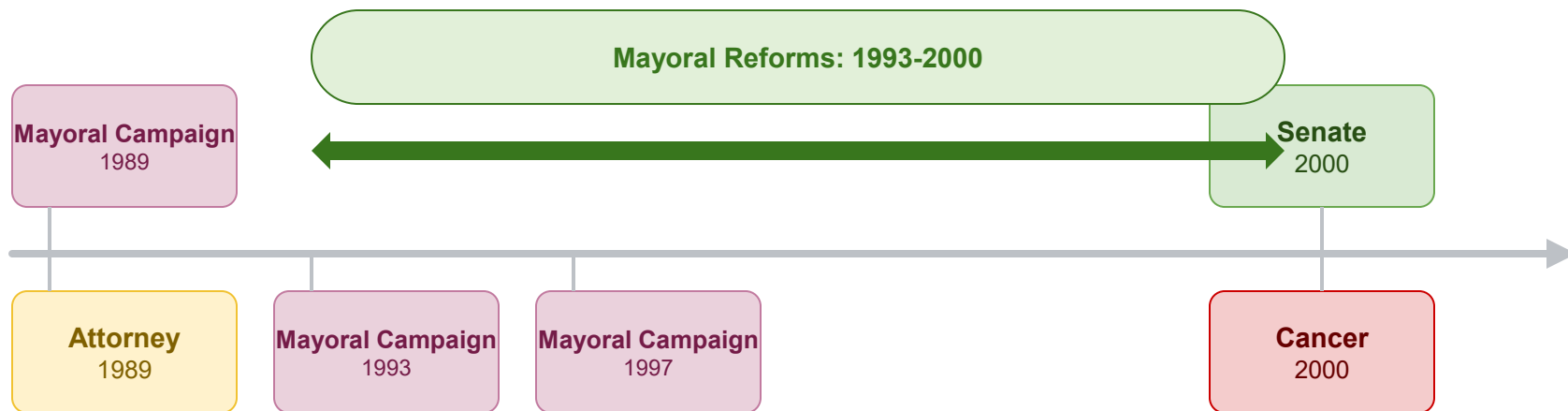
(Subtopics)

Different facets or topics covered within the retrieved documents that evolve over time.

Time

The temporal dimension that defines when specific aspects were most prominent or significant.

Temporal Diversification



Query: Rudi Giuliani

- Aspects are diverse across time
- Time periods are aspect diverse

2-Dimensional Coverage Problem

Diversification based approaches

[Singh, Nejdl, Anand., 2016]