

A Language Modeling Approach for Temporal Information Needs

Klaus Berberich, Srikanta Bedathur
Omar Alonso, Gerhard Weikum

2010, March 29th

ECIR 2010 – Milton Keynes



Motivation

- Information needs having a **temporal dimension**, e.g.:
 - FIFA World Cup tournaments of the 1990's*
 - Movies that won an Academy Award in 2007*
 - Crusades of the 12th century*
- Queries containing a temporal expression (e.g., **in 1998**)
 - **indicate** an underlying temporal information need
 - make up **1.5% of general web queries** (Nunes et al. '08)
 - are more common for **specific domains** (e.g., News or Sports) and/or **specific user groups** (e.g., journalists or historians)
- **But**: Not well-supported by existing retrieval models!

Outline

- Motivation
- Temporal Expressions
- Language Models for Temporal Information Needs
- Experimental Evaluation
- Conclusion



Temporal Expressions

- Temporal expressions are **frequent across many classes of documents** and can be categorized as:
 - explicit (e.g., **March 29th, 2010** or **September 1872**)
 - implicit (e.g., **Christmas 2009** or **New Year's Eve 2000**)
 - relative (e.g., **yesterday, last month, or in January**)
- **Off-the-shelf tools** available to identify and interpret temporal expressions (e.g., **TARSQI** and **TimexTag**)
- **TimeML** mark-up language specification to annotate temporal expressions found in a document



Challenges

- Existing retrieval models **ignore temporal expressions and their meaning** and therefore fail to match, e.g.:

fifa world cup
1990's



During the 90's the FIFA World Cup was won by Germany (*in 1990*), Brazil (*in 1994*), and France (*in 1998*)...

- Meaning** of a temporal expression is often **uncertain**, e.g.:

France won the FIFA World Cup *in 1998*

In 1998 Bill Clinton was President of the U.S.

Nagano hosted the Winter Olympics *in 1998*

Temporal Expression Model

- We **formally model** temporal expressions as 4-tuples:

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

that record the **earliest/latest begin/end of time intervals** that T may refer to

- The temporal expression T **may thus refer to any time interval** $[b, e]$ such that

$$b \leq e \wedge tb_l \leq b \leq tb_u \wedge te_l \leq e \leq te_u$$

- **In 1998**, e.g, is represented as
(1998/01/01, 1998/12/31, 1998/01/01, 1998/12/31)

Outline

- Motivation
- Temporal Expressions
- **Language Models for Temporal Information Needs**
- Experimental Evaluation
- Conclusion



Document & Query Model

- We distinguish between the **textual part** and **temporal part** of a **document**
 - d_{text} is a bag of textual terms
 - d_{time} is a bag of temporal expressions
- We distinguish between the **textual part** and **temporal part** of a **query**
 - q_{text} is a bag of textual terms
 - q_{time} is a bag of temporal expressions
- Two modes of deriving a query from the user's input
 - **inclusive mode**: q_{text} includes all input terms
 - **exclusive mode**: q_{text} excludes input terms that are part of a temporal expression



Language Model Framework

- **Query-likelihood approach** assuming **independent generation** of textual and temporal query part

$$P(q | d) = P(q_{\text{text}} | d_{\text{text}}) \times P(q_{\text{time}} | d_{\text{time}})$$

- $P(q_{\text{text}} | d_{\text{text}})$ estimated using a **unigram language model with Jelinek-Mercer smoothing** as

$$\prod_{q \in q_{\text{text}}} \lambda \cdot P(q | d_{\text{text}}) + (1 - \lambda) \cdot P(q | C)$$

Language Model Framework

- We assume that **query temporal expressions** are **generated independently** from each other

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

- **Two-step generation** of query temporal expression Q

(I) Draw a temporal expression T at **uniform random**

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

(II) Generate Q from T

Uncertainty-ignorant Approach

- The temporal expression T **only generates itself**

$$P(Q | T) = \begin{cases} 1 & : Q = T \\ 0 & : \text{otherwise} \end{cases}$$

- **Ignores** temporal expressions' **inherent uncertainty**
- **Profits** from the extraction of temporal expressions

fifa world cup
| 1990's



*During the **90's** the FIFA World Cup was won by Germany (**in 1990**), Brazil (**in 1994**), and France (**in 1998**)...*

Uncertainty-aware Approach

- We assume that **any time interval** $[q_b, q_e]$ that the query temporal expression Q may refer to **is equally likely**

$$P(Q | T) = \frac{1}{|Q|} \sum_{[q_b, q_e] \in Q} P([q_b, q_e] | T)$$

- The temporal expression T generates **any time interval** $[q_b, q_e]$ **that it may refer to with equal probability**

$$P([q_b, q_e] | T) = \begin{cases} 1/|T| & : [q_b, q_e] \in T \\ 0 & : \text{otherwise} \end{cases}$$

Uncertainty-aware Approach

- Intuitively, $P(Q | T)$ reflects the probability that the **user issuing the query** and **the author writing the document** had the same time interval in mind
- The definition can be **simplified** as

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

treating Q and T as sets of time intervals

- $P(Q | T)$ **can be computed efficiently** without “materializing” the huge but finite sets of time intervals

Uncertainty-aware Approach

- **Considers** temporal expressions' **inherent uncertainty**
- **Profits** from the extraction of temporal expressions

fifa world cup
| 1990's



*During the **90's** the FIFA World Cup was won by Germany (**in 1990**), Brazil (**in 1994**), and France (**in 1998**)...*

Outline

- Motivation
- Temporal Expressions
- Language Models for Temporal Information Needs
- **Experimental Evaluation**
- **Conclusion**



We are interested in exploring search scenarios where **temporal information** is important to satisfy an information need. By temporal information we mean any time reference (e.g., “August 1999”, “20th century”, or “January 1 2002”).

Instructions

You're given an incomplete query consisting only of a **person, location, or organization name**. Please complete the query by adding a **time reference** that you think fits the person, location, or organization.

Examples

- Given **barack obama**, you could add *January 20, 2009* as the day of Barack Obama's inauguration
- Given **lehman brothers**, you could add *September 2008* as the month when the company went bankrupt
- Given **beijing**, you could add *2008* as the year when the Olympic Games took place in Beijing
- Given **samuel adams**, you could add *1790s* as the decade when Samuel Adams was Governor of Massachusetts
- Given **red cross**, you could add *1863* as the year when the Red Cross was founded
- Given **germany**, you could add *October 3, 1990* as the day of the German Reunification

Task

Please complete the following query by adding a **time reference** that fits the given **person, location, or organization name**.

boston red sox

We appreciate your comments (e.g., why you picked this particular time reference)!

Submit

Datasets & Methods

- Document collections:
 - New York Times Annotated Corpus
 - English Wikipedia (as of July '09)
- Temporal expressions annotated using **TARSQI**
- **40 queries** collected using **Amazon Mechanical Turk**

boston red sox [october 27, 2004]

pink floyd [march 1973]

michael jordan [1990s]

soccer [21st century]

voyager [september 5, 1977]

poland [december 1970]

sewing machine [1850s]

...

kurt cobain [april 5, 1994]

babe ruth [1921]

mickey mouse [1930s]

jazz music [21st century]

berlin [october 27, 1961]

wright brothers [1905]

siemens [19th century]

Relevance Assessments & Measures

- **Relevance assessments** on pooled query results collected using **Amazon Mechanical Turk**
 - **binary** relevance assessments with “I don’t know” option
 - **mandatory justification** of relevance assessment
 - **three assessors** per query-document pair
- **Fleiss’ Kappa** statistic indicates **fair level of agreement**
 - New York Times Annotated Corpus (0.36)
 - English Wikipedia (0.40)
- We measure **retrieval effectiveness** using
 - Precision @ 10 (P@10)
 - Normalized Discounted Cumulative Gain @ 10 (nDCG@10)

- A document is considered relevant if it contains **both textual and temporal information matching the query**
- Only work with **meaningful explanations** will be accepted (i.e., do not just write "relevant" or "not relevant")

Task

Please judge the relevance of the following document to the query **musket 16th century**. Remember, a document is considered relevant if it contains **both textual and temporal information** matching the query.



The screenshot shows a Wikipedia article page for "Pike and shot". At the top, there is a banner for donations and a "Log in / create account" link. Below the banner are navigation tabs for "article", "discussion", "edit this page", and "history". The article title "Pike and shot" is prominently displayed. Below the title, it says "From Wikipedia, the free encyclopedia". A grey box contains a warning: "This is an old revision of this page, as edited by Ingolfson (talk | contribs) at 06:18, 4 July 2009. It may differ significantly from the current revision." Below this warning are navigation links: "(diff) ← Previous revision | Current revision (diff) | Newer revision → (diff)". The main text of the article begins with "Pike and shot is a historical method of infantry combat, and". On the left side, there is a "WIKIPEDIA The Free Encyclopedia" logo and a "navigation" menu with a link to "Main page".

Please judge the relevance of the above document to the query **musket 16th century** as follows.

- Relevant.** A relevant document containing both textual and temporal information relevant to the query.
- Not relevant.** The document is not good because it doesn't contain any relevant information.
- I don't know.** I don't have enough information to evaluate this document.

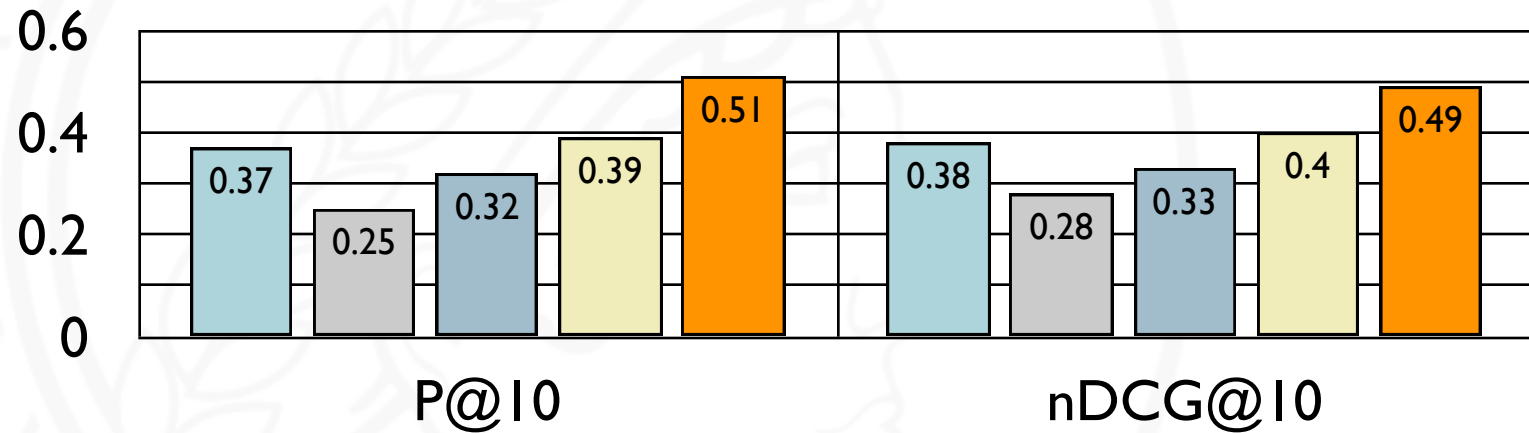
Please explain why you think the document is relevant or not relevant!

Submit

Relevance Assessments & Measures

- **Relevance assessments** on pooled query results collected using **Amazon Mechanical Turk**
 - **binary** relevance assessments with “I don’t know” option
 - **mandatory justification** of relevance assessment
 - **three assessors** per query-document pair
- **Fleiss’ Kappa** statistic indicates **fair level of agreement**
 - New York Times Annotated Corpus (0.36)
 - English Wikipedia (0.40)
- We measure **retrieval effectiveness** using
 - Precision @ 10 (P@10)
 - Normalized Discounted Cumulative Gain @ 10 (nDCG@10)

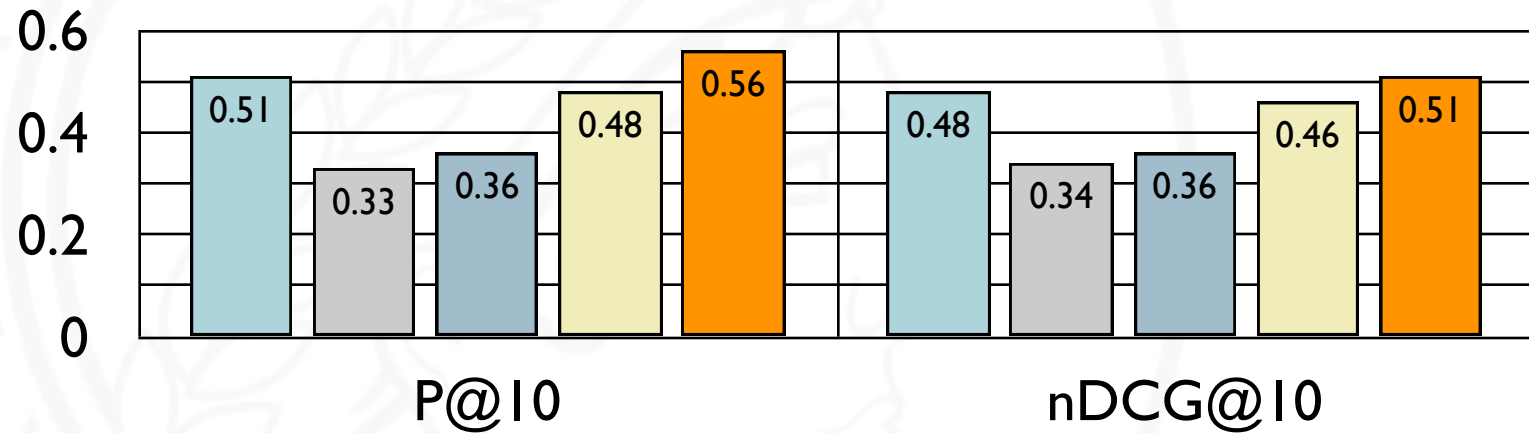
Retrieval Effectiveness Overall



New York Times



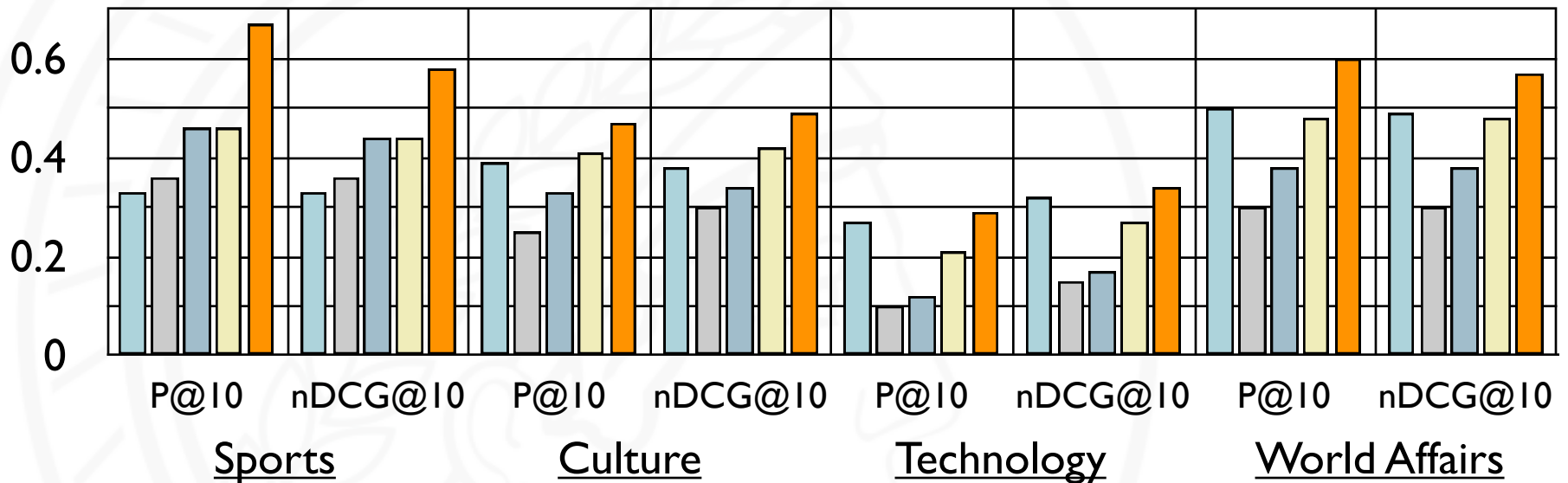
Retrieval Effectiveness Overall



Wikipedia



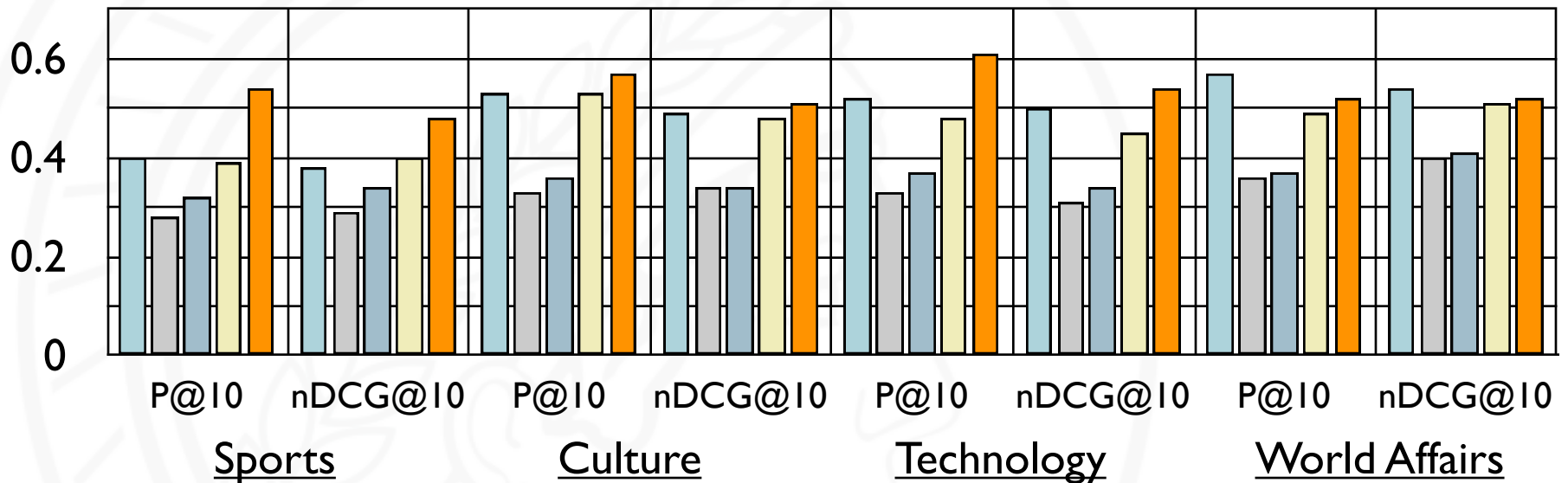
Retrieval Effectiveness by Topic



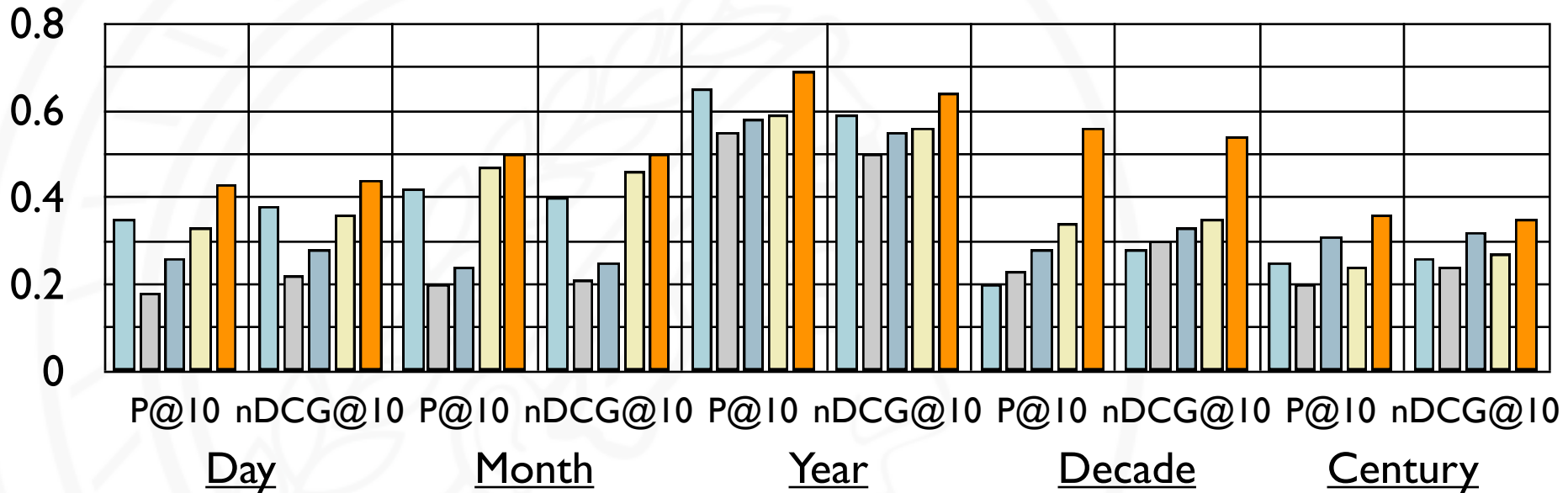
New York Times



Retrieval Effectiveness by Topic



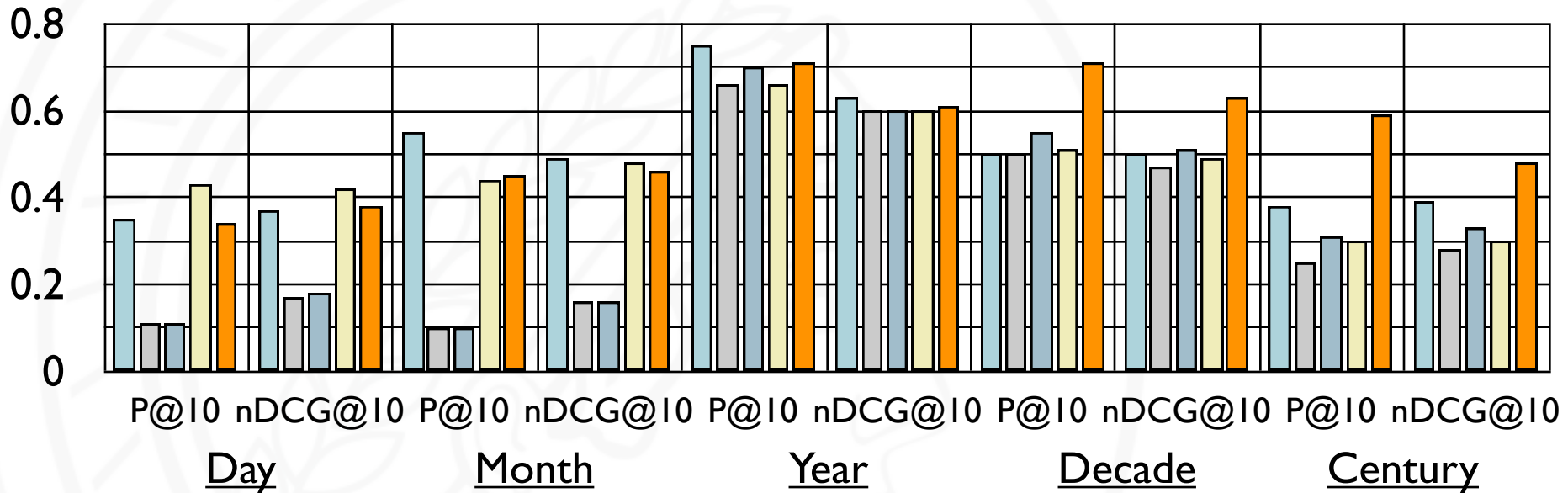
Retrieval Effectiveness by Granularity



New York Times



Retrieval Effectiveness by Granularity



Wikipedia



Outline

- Motivation
- Temporal Expressions
- Language Models for Temporal Information Needs
- Experimental Evaluation
- Conclusion



Conclusion

- Our approach **integrates temporal expressions seamlessly** into a language modeling framework
- Experiments show that **temporal expressions are helpful to better satisfy temporal information needs** if their **inherent uncertainty is taken into account**
- More (details on) experiments and links to download **extracted temporal expressions** and **relevance assessments** available in our technical report!





Thanks! Questions?

<http://www.mpi-inf.mpg.de/~kberberi/ecir2010/>

