GTE: A Distributional Second-Order Co-Occurrence Approach to Improve the Identification of Top Relevant Dates in Web Snippets

Ricardo Campos¹, 4, 6, Gaël Dias², 6, Alípio Jorge³, 4, Célia Nunes⁵, 6

¹Tomar Polytechnic Institute, Tomar, Portugal
²HULTEC/GREYC, University of Caen Basse-Normandie, France
³Faculty of Sciences, University of Porto, Porto, Portugal
⁴LIAAD-INESC Porto L.A, Porto, Portugal
⁵Department of Mathematics, University of Beira Interior, Covilhã, Portugal
⁶Center of Mathematics, University of Beira Interior, Covilhã, Portugal

Explicit Temporal Queries

Related work in T-IR shows us that most of the existing approaches are built under the **assumption** that **users** always **supply some form of temporal context**: 

![FIFA World Cup 2014 logo](image-url)
Implicit Temporal Queries

This is not always the case:

FIFA World Cup Germany

2006
Understand the Nature of Implicit Temporal Queries

Understanding the temporal nature of this type of queries, is therefore of the utmost importance.
State-of-the-art

- In this context, most state-of-the-art methodologies consider any occurrence of temporal expressions as equally relevant to an implicit temporal query:

<table>
<thead>
<tr>
<th>Title</th>
<th>2011 Haiti Earthquake Anniversary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snippet</td>
<td>As of 2010 (see 1500 photos), the following major earthquakes have been recorded in Haiti. The 1st one occurred in 1564. 2010 has been a tragic date, however in 2012 Haiti will organize the Carnival...</td>
</tr>
</tbody>
</table>

Understanding the temporal nature of this type of queries, is therefore of the utmost importance.
In this context, most state-of-the-art methodologies consider any occurrence of temporal expressions as equally relevant to an implicit temporal query. This is obviously not true.

<table>
<thead>
<tr>
<th>Title</th>
<th>2011 Haiti Earthquake Anniversary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snippet</td>
<td>As of 2010 (see 1500 photos), the following major earthquakes have been recorded in Haiti. The 1st one occurred in 1564. 2010 has been a tragic date, however in 2012 Haiti will organize the Carnival...</td>
</tr>
</tbody>
</table>
Set Goals

- Define the **temporal** intents of any given implicit **query**;

Hence our goal is twofold

1. Select the most relevant dates for a given query

As referred by Berberich et al (2010) this is an interesting future research direction for which there isn’t yet a clear solution.

[www.linkedin.com/in/camposricardo]

Haiti earthquake

1564  2010  2011
Set Goals

- Define the **temporal** intents of any given implicit **query**;

  Hence our goal is twofold

  1. Select the most relevant dates for a given query
  2. Discard all irrelevant or incorrect ones.

As referred by Berberich et al (2010) this is an interesting future research direction for which there isn’t yet a clear solution.
We propose a novel second-order similarity measure to assess the temporal similarity between a query and a date;

We exhaustively evaluate our measure on a real-world dataset;

We publicly provide a gold-standard, previously unavailable.
Outline

- Approach
- Evaluation
- Conclusions

[www.linkedin.com/in/camposricardo]

[w w w . c c c . i p t . p t / ~ r i c a r d o]
System Architecture

Web Search and Web Snippet Module

GTE: Generic Temporal Similarity Measure

Temporal Classification Module

- Threshold-based classification
- SVM classifier
Given a query $q$ issued by a user, e.g., $q = \{\text{haiti earthquake}\}$

We obtain a collection of web snippets $S = \{S_1, S_2, \ldots, S_n\}$
Select Best Words ($W_s$) and Candidate Dates ($D_s$)

- Let $W_s = \{w_1, w_2, \ldots, w_k\}$ be the set of $k$ distinct best relevant words/multi-words extracted for the query $q$ within the set of web snippets $S$:

$$W_s = \{\text{haiti earthquake; major earthquakes; haiti; catastrophic damage; Port-au-Prince; Concepción de la Vega}\}$$

- Let $D_s = \{d_1, d_2, \ldots, d_t\}$ be the set of $t$ distinct candidate dates retrieved from the set of web snippets $S$ returned for the query $q$:

$$D = \{1500; 1564; 2010; 2011\}$$
Problem Definition

Given a query $q$ and a date $d_i$ assign a degree of relevance to each $(q, d_i)$ pair.

To model this relevance, we define a temporal similarity value $v$ given by a similarity measure $sim$.

$$v = sim(q, d_i), \; v \in [0,1].$$

- The proposed formulation tries to identify relevant dates $d_i$ for $q$;
- **Minimize any errors** that might arise from considering irrelevant or wrong dates.

[www.linkedin.com/in/camposricardo] [www.ccc.ipt.pt/~ricardo]
Problem Definition

According to our investigation, the relevance between a \((q, d_i)\) pair is better defined if:

- Instead of just focusing on the self-similarity between \(q \leftrightarrow d_i\)

---

[Haiti Earthquake]

<table>
<thead>
<tr>
<th>1500</th>
<th>1564</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>
According to our investigation, the relevance between a \((q, d_i)\) pair is better defined if

- We compute the similarities between \(W^* d_i\)

### Problem Definition

<table>
<thead>
<tr>
<th>Haiti Earthquake</th>
<th>1500</th>
<th>1564</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>major earthquakes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>haiti</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catastrophic damage</td>
<td>0</td>
<td>0</td>
<td>v</td>
<td>0</td>
</tr>
<tr>
<td>port-au-prince</td>
<td>0</td>
<td>0</td>
<td>v</td>
<td>0</td>
</tr>
<tr>
<td>concepción de la vega</td>
<td>0</td>
<td>v</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

[www.linkedin.com/in/camposricardo] [www.ccc.ipt.pt/~ricardo]
In order to compute all these values, we propose a new generic temporal similarity measure called:

\[ \text{GenTempEval}(q, d_i) = F(\text{sim}(X, d_i)), X \in W^* \]

\(F\) is an aggregation function that combines the different similarity values \(\text{sim}(X, d_i)\) in a single one;

For this purpose, we consider three different \(F\) functions:

- Arithmetic Mean;
- Median;
- Max / Min;
GenTempEval

In order to compute all these values, we propose a new generic temporal similarity measure called:

\[ GenTempEval(q, d_i) = \text{Median}(\text{sim}(X, d_i)), \ X \in W^* \]

For the \textit{sim} measure we apply the \textit{InfoSimba} (see Dias et al.) similarity measure, a semantic vector space model supported by corpus-based correlations:

Each word \( X \) and date \( d_i \) is defined in terms of a context vector

\[
\begin{array}{c}
X \\
d_i
\end{array}
\]

\[
\begin{array}{c}
\hspace{1cm} \\
\hspace{1cm}
\end{array}
\]

\[
\begin{array}{c}
\hspace{1cm} \\
\hspace{1cm}
\end{array}
\]

\[
\begin{array}{c}
\ldots \\
\ldots
\end{array}
\]
For this purpose, **five possible representations** have been defined:

- \((q; d_i)\) Contextual Vectors Representation

\[ X \text{ Word Word Word …} \]

\[ d_i \text{ Word Word Word …} \]
(q,d_i) Contextual Vectors Representation

For this purpose, **five possible representations** have been defined:

\[
\begin{array}{cccccc}
X & \text{Date} & \text{Date} & \text{Date} & \cdots \\
\text{d}_i & \text{Date} & \text{Date} & \text{Date} & \cdots \\
\end{array}
\]

\( (D;D) \)
For this purpose, **five possible representations** have been defined:

\[(q,d_i)\] Contextual Vectors Representation

- (W;D)
- \(X\) Word Word Word \(\ldots\)
- \(d_i\) Date Date Date \(\ldots\)
### (q,d_i) Contextual Vectors Representation

For this purpose, **five possible representations** have been defined:

- $(q,d_i)$

<table>
<thead>
<tr>
<th>X</th>
<th>Date</th>
<th>Date</th>
<th>Date</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i$</td>
<td>Word</td>
<td>Word</td>
<td>Word</td>
<td>...</td>
</tr>
</tbody>
</table>

$\bullet$ $(D;W)$

---

[w w w. l i n k e d i n . c o m / i n / c a m p o s r i c a r d o]

[w w w. c c c . i p t . p t / ~ r i c a r d o]
For this purpose, **five possible representations** have been defined:

For the query/document pair $$(q,d_i)$$, we consider two types of contextual vectors representation:

- **X**: A representation where the query and document are aligned, i.e., $$(X;WD;WD)$$.
- **d_i**: A representation where the document is used as a context for the query, i.e., $$(d_i;WD;WD)$$.

These representations aim to capture the temporal relationships between words and dates within web snippets.
\[ IS(V_x, V_y) = \frac{\sum_{i \in V_x} \sum_{j \in V_y} S(i, j)}{\left( \sum_{i \in V_x} \sum_{j \in V_x} S(i, j) + \sum_{i \in V_y} \sum_{j \in V_y} S(i, j) - \sum_{i \in V_x} \sum_{j \in V_y} S(i, j) \right)} \]

where \( V_x \) and \( V_y \) are the context vectors of \( X \) and \( d_i \) respectively.

The similarity between each pair \( S(i,j) \) of the two context vectors is determined by any **first order similarity measure** \( S(.,.) \), e.g., \((PMI, EI \text{ or } DICE)\)

This requires the definition of \( M_{ct} \)
**M_{ct}: Conceptual Temporal Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Haiti Earthquake</th>
<th>major earthquakes</th>
<th>haiti</th>
<th>catastrophic damage</th>
<th>port-au-prince</th>
<th>concepción de la vega</th>
<th>1500</th>
<th>1564</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haiti Earthquake</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>major earthquakes</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>haiti</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>catastrophic damage</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>port-au-prince</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>concepción de la vega</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>1500</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>1564</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>2010</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
<tr>
<td>2011</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
<td>·</td>
</tr>
</tbody>
</table>
Two Different Strategies

- **Threshold-based classification**
  Given a \((q,d_i)\) pair the system automatically classifies a date based on the following:
  - Relevant, if \(GTE(q,d_i) \geq \lambda\)
  - Irrelevant, if \(GTE(q,d_i) < \lambda\)

- **SVM Learning Model**

<table>
<thead>
<tr>
<th>Query</th>
<th>(d_i)</th>
<th>Class</th>
<th>IS</th>
<th>DICE</th>
<th>PMI</th>
<th>EI</th>
<th>TF.IDF</th>
<th>WebDICE</th>
<th>WebJaccard</th>
<th>WebPMI</th>
<th>..........</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haiti Earthquake</td>
<td>1500</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Haiti Earthquake</td>
<td>1564</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Haiti Earthquake</td>
<td>2010</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Haiti Earthquake</td>
<td>2011</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

[www.linkedin.com/in/camposricardo] [www.ccc.ipt.pt/~ricardo]
Test Queries

42 representative clear-concept implicit temporal queries:

- non-ambiguous in concept;
- temporal in purpose.

[w w w. l i n k e d i n . c o m / i n / c a m p o s r i c a r d o]  [w w w. c c c. i p t . p t / ~ r i c a r d o]
Data Description

Since no benchmark for \((q,d_i)\) pairs exists, we built a new web-based dataset;

We queried the Bing search engine for each of the 42 queries, collecting the 50 best relevant web results;

- **582** relevant web snippets with **years**;
- **235** distinct \((q,d_i)\) pairs;

Each one was assigned a **relevance label** by a human judge on a **2-level scale**:

- not a date or irrelevant (score 0);
- relevant date (score 1).

<table>
<thead>
<tr>
<th>Score</th>
<th># ((q,d_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>86</td>
</tr>
<tr>
<td>1</td>
<td>149</td>
</tr>
</tbody>
</table>

[www.linkedin.com/in/camposricardo] [www.ccc.ipt.pt/~ricardo]
Our Approach

Several **different versions of the GenTempEval** combined with **InfoSimba** were considered:

They are represented as $IS_{(X;Y)}_{SM_F}$
Our Approach

Several different versions of the GenTempEval combined with InfoSimba were considered:

They are represented as $IS_{(X;Y)_{SM_F}}$

- $IS \rightarrow$ InfoSimba;
Our Approach

Several different versions of the *GenTempEval* combined with *InfoSimba* were considered:

They are represented as $IS_{(X;Y)}_{SM,F}$

- $IS \rightarrow$ InfoSimba;
- $(X;Y) \rightarrow$ means the type of contextual vector;
Several **different versions of the GenTempEval** combined with **InfoSimba** were considered:

They are represented as $IS_{(X;Y)}_{SM-F}$

- **IS** $\rightarrow$ InfoSimba;
- **(X;Y)** $\rightarrow$ means the type of contextual vector;
- **SM** $\rightarrow$ any similarity measure of first order used with IS (e.g., PMI, DICE,...);
Several different versions of the GenTempEval combined with InfoSimba were considered:

They are represented as $IS_{(X;Y)}_{SM\_F}$ e.g. $IS_{(WD;WD)}_{DICE\_Median}$

- **IS** → InfoSimba;
- **(X;Y)** → means the type of contextual vector;
- **SM** → any similarity measure of first order used with IS (e.g., PMI, DICE,…);
- **F** → the aggregator function.

Further experiments have been performed based on the InfoSimba measure without the use of any paradigm. Overall, all of these measures are denoted $IS_{(X;Y)}_{SM}$. 
Baseline Methods

- PMI;
- DICE;
- Jaccard;
- SCP;
- NGoogleDistance;
- WebJaccard;
- WebOverlap;
- WebDICE;
- WebPMI;

Without the aggregator function, denoted SM
# Baseline Methods

With the aggregator function, denoted $SM_F$

- PMI$_F$
- DICE$_F$
- Jaccard$_F$
- SCP$_F$
- NGoogleDistance$_F$
- WebJaccard$_F$
- WebOverlap$_F$
- WebDICE$_F$
- WebPMI$_F$
For each of the contextual vectors representations, we have performed a set of experiments with different sizes, $N$.

<table>
<thead>
<tr>
<th>Setting Evaluation Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
</tr>
<tr>
<td>$d_i$</td>
</tr>
</tbody>
</table>

$N = ?$
For each of the contextual vectors representations, we have performed a set of experiments with different sizes, $N$.
For each of the contextual vectors representations, we have performed a set of experiments with different sizes, $N$, and threshold values, $T$. 

$$x \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$d_i \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$x \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$d_i \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$N = ?$$

$$\ldots \rightarrow T$$
To compare the different measures, **we used the point biserial correlation coefficient**, a statistical measure of the correlation between one continuous and one binary variable.

<table>
<thead>
<tr>
<th>(q,d_i)</th>
<th>Class</th>
<th>GenTempEval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar Movie - 2009</td>
<td>1</td>
<td>0.670</td>
</tr>
<tr>
<td>Avatar Movie - 2011</td>
<td>0</td>
<td>0.346</td>
</tr>
<tr>
<td>Bp Oil Spill - 2010</td>
<td>1</td>
<td>0.838</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Point Biserial Correlation</td>
<td>-</td>
<td>0.800</td>
</tr>
</tbody>
</table>
We developed a ROC curve. The red line indicates an almost perfect classifier with an Area Under Curve (AUC) of 0.960, for the IS_(WD;WD)_DICE_Median, onwards denoted BGTE.
When compared to the baseline measures, BGTE reaches the highest values:
Threshold-based Classification

In order to assess the temporal similarity between a query and a date, i.e., to determine whether a date is relevant or not for a given query.

We use a classical threshold-based strategy:

Given a \((q, d_i)\) pair:

- **Relevant**, if \(GTE(q, d_i) \geq \lambda\)
- **Irrelevant**, if \(GTE(q, d_i) < \lambda\)

<table>
<thead>
<tr>
<th>Recall</th>
<th>Precision</th>
<th>Accuracy</th>
<th>Balanced Accuracy</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.919</td>
<td>0.944</td>
<td>0.914</td>
<td>0.913</td>
<td>0.931</td>
</tr>
</tbody>
</table>

[w w w. li n k e d i n. c o m / i n / c a m p o s r i c a r d o ]

[w w w. c c c. i p t. p t / ~ r i c a r d o]
## SVM Learning Model

<table>
<thead>
<tr>
<th>Attribute Set</th>
<th>Balanced Accuracy</th>
<th>Average F1</th>
<th>Average AUC</th>
<th>Correct Date</th>
<th>Incorrect or Irrelevant Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>All Measures</td>
<td>0.903</td>
<td>0.902</td>
<td>0.894</td>
<td>0.920</td>
<td>0.926</td>
</tr>
<tr>
<td>All Measures after Feature Selection</td>
<td>0.886</td>
<td>0.885</td>
<td>0.876</td>
<td>0.907</td>
<td>0.913</td>
</tr>
</tbody>
</table>

- Feature selection may not lead to improved results;

- Compared to the threshold-based classification strategy, the results obtained by the SVM classification are worse than only using BGTE.
Comparison of BGTE against a Query Log approach

<table>
<thead>
<tr>
<th>query</th>
<th>Google_QLogs</th>
<th>Yahoo_QLogs</th>
<th>BGTE</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.653</td>
<td>0.647</td>
<td>0.748</td>
<td>0.634</td>
</tr>
<tr>
<td>Recall</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F1</td>
<td>0.790</td>
<td>0.786</td>
<td>0.856</td>
<td>0.776</td>
</tr>
</tbody>
</table>

- Results show that query logs are able to return a great number of potential query related year dates, when compared to Web snippets;

- But, interestingly, we found that a large number of these temporally explicit queries consist of misleading temporal relations i.e. users may execute incorrect temporal queries as they may not know the exact date related to their query;

[w w w. l i n k e d i n . c o m / i n / c a m p o s r i c a r d o] [w w w. c c c . i p t . p t / ~ r i c a r d o]
We proposed a new temporal similarity measure, the \textit{GenTempEval}, in order to compute the temporal intent(s) of \((q,d_i)\) pairs;

We showed that the combination of the second order similarity measure \textit{InfoSimba} with the \textit{DICE} coefficient, a representation of words and dates and the \textit{Median} aggregator function shows better results than all other combinations;

We believe that the proposed method will be useful to disambiguate a large class of implicit temporal queries.
Thanks for your attention!

Both experimental datasets are available for download at www.ccc.ipt.pt/~ricardo/software

Polytechnic Institute of Tomar is online at http://www.ipt.pt
LIAAD is online at http://liaad.up.pt
CMAT is online at http://www.cmatubi.ubi.pt/english/
University of Caen is online at http://www.unicaen.fr/

Ricardo Campos is online at www.ccc.ipt.pt/~ricardo
Gaël Dias is online at https://dias.users.greyc.fr/?
Alípio Jorge is online at http://www.liaad.up.pt/~amjorge/
Célia Nunes is online at http://www.mat.ubi.pt/~celia/