Enriching Temporal Query Understanding through Date Identification: How to Tag Implicit Temporal Queries?

Ricardo Campos 1, 4, Gaël Dias 2, Alípio Jorge 3, 4, Célia Nunes 5

1 Tomar Polytechnic Institute, Tomar, Portugal
2 HULTEC/GREYC, University of Caen Basse-Normandie, France
3 Faculty of Sciences, University of Porto, Porto, Portugal
4 LIAAD-INESC Porto L.A, Porto, Portugal
5 Department of Mathematics and Center 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:
Implicit Temporal Queries

This is not always the case:

FIFA World Cup Germany

2006
Implicit Temporal Queries

This is not always the case:

**George Bush Iraq War**

1991
Implicit Temporal Queries

This is not always the case:

George Bush Iraq War

2003
Understanding the temporal nature of this type of queries, is therefore of the utmost importance.
INTRODUCTION

Improving Search Results Exploration

- Provide Temporal query expansion;

Understanding the temporal nature of this type of queries is therefore of the utmost importance in order to improve search results exploration.

[www.linkedin.com/in/camposricardo]

[Ricardo Campos, Gaël Dias, Alípio Jorge, Célia Nunes]
Understanding the temporal nature of this type of queries, is therefore of the utmost importance in order to improve search results exploration.

MOTIVATION

Contributions

• Temporal query expansion;

• Promote the use of Timelines;
INTRODUCTION

Problem Formulation

Evaluation

Approach

CONCLUSIONS

MOTIVATION

Contributions

Improve Search Results Exploration

- **Temporal Query Expansion**;
- **Timelines**;
- Perform **Temporal Clustering** of the results;

Understanding the temporal nature of this type of queries, is therefore of the utmost importance in order to improve search results exploration.
• Nevertheless, there is very little work addressing this particular issue;

• Moreover, none of the works, extract temporal information from the contents of the document.

<table>
<thead>
<tr>
<th>Work</th>
<th>Approach</th>
<th>Collection of Temporal Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dakka et. al</td>
<td>Metadata</td>
<td>Web News Articles</td>
</tr>
<tr>
<td>Jones et. al</td>
<td>Metadata</td>
<td>Web News Articles</td>
</tr>
<tr>
<td>Kanhabua et. al</td>
<td>Metadata</td>
<td>Web News Articles</td>
</tr>
</tbody>
</table>
Related Work

- Nevertheless, there is very **little work addressing this particular issue**;

- Moreover, **none of the works** extract temporal information from the contents of the document.

<table>
<thead>
<tr>
<th>Work</th>
<th>Approach</th>
<th>Collection of Temporal Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metzler et. al</td>
<td>Usage</td>
<td>Query Logs</td>
</tr>
</tbody>
</table>
Understand the temporal nature of any 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.
INTRODUCTION

Problem Formulation

Evaluation

Conclusions

Motivation

CONTRIBUTIONS

Set Goals

- **Understand** the *temporal nature* of any given *query*;

- Based on *web content* analysis rather than following a metadata-based approach.

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

Problem Formulation

Approach

Evaluation

Motivation

Conclusions

CONTRIBUTIONS

Set Goals

- **Understand** the *temporal nature* of any given *query*;

- Based on **web content** analysis rather than following a metadata-based approach.

- Specifically, we aim at **dating implicit temporal queries**;

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

[www.ccc.ipt.pt/~ricardo]
As referred by Berberich et al (2010) this is an interesting future research direction for which there isn’t yet a clear solution.

- **Understand** the **temporal nature** of any given **query**;

- Based on **web content** analysis rather than following a metadata-based approach.

- Specifically, we aim at **dating implicit temporal queries**;

**Haiti earthquake**

[www.linkedin.com/in/camposricardo]

1564  
2010  
2011
1. **Firstly**, how can we ensure that the detected **number is actually a date** and not a single number (e.g., archive of 1000's of romantic love poems);

2. **Secondly**, how can we **ensure** with a strong confidence, the proper **linkage** between the **dates** and the **concepts** found in the document:

   Miss Universe was held this year in Bahamas. **2008 was an incredible year, but everybody is waiting for the FIFA South Africa Football World Cup. From** [www.nbc.com/wc2010](http://www.nbc.com/wc2010)
Contributions

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.

[www.linkedin.com/in/camposricardo]

[www.ccc.ipt.pt/~ricardo]
Outline

- Problem Formulation
- Approach
- Evaluation
- Conclusions
Step 1: Query Formulation

- Implicit Temporal Queries
- Web Snippet Collection
Step 2: Document Annotation

- Word / Multi-Word Extraction
- Temporal Expression Extraction
- Model the relations between encountered timestamps and words in the web snippets
Step 3: Temporal Query Disambiguation

- For each d
  Compute Sim \( (q,d) \)

- Output: \( (q,d) \) relevance
  Tag query with relevant years
  Filter out irrelevant ones

Diagram:

- Query Formulation
  - Web Snippets Collection
  - Implicit Temporal Queries

- Document Annotation
  - Word / Multi-Word Extraction
  - Temporal Expression Extraction

- Temporal Query Disambiguation
  - Compute Sim \( (q,d) \)
  - Output: \( (q,d) \) relevance
Information Extraction

- Given a query $q$ issued by a user, e.g., $q = \{\text{haiti earthquake}\}$
- We obtain a collection of web snippets $S = \{S_1, S_2, \ldots, S_n\}$

<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 here), the following major earthquakes have been recorded in Haiti. The first one occurred in 1564.</td>
</tr>
<tr>
<td>$W_{S_1}$</td>
<td>haiti earthquake; major earthquakes; haiti</td>
</tr>
<tr>
<td>$D_{S_1}$</td>
<td>1500; 1564; 2010; 2011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title</th>
<th>Haiti Earthquake Relief</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snippet</td>
<td>On January 12, 2010, a massive earthquake struck the nation of Haiti, causing catastrophic damage inside and around the capital city of Port-au-Prince.</td>
</tr>
<tr>
<td>$W_{S_2}$</td>
<td>haiti earthquake; haiti; catastrophic damage; Port-au-Prince</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title</th>
<th>Haiti Earthquake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snippet</td>
<td>The first great earthquake mentioned in histories of Haiti occurred in 1564 in what was still the Spanish colony. It destroyed Concepción de la Vega.</td>
</tr>
<tr>
<td>$W_{S_3}$</td>
<td>haiti earthquake; haiti; Concepción de la Vega</td>
</tr>
<tr>
<td>$D_{S_3}$</td>
<td>1564</td>
</tr>
</tbody>
</table>
Let $W = \{w_1, w_2, \ldots, w_p\}$ be the set of distinct best relevant words/multi-words extracted for the query $q$ within the set of web snippets $S$:

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

Let $D = \{d_1, d_2, \ldots, d_m\}$ be the set of temporal patterns retrieved from the set of web snippets $S$ returned for the query $q$:

$$D = \{1500; 1564; 2010; 2011\}$$
Given a query $q$ issued by a user, and the set of dates $D$ retrieved from the set of web snippets $S$ returned for the query $q$, 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 $\text{sim}$.

\[ v = \text{sim}(q, d_i), \forall v \in [0,1]. \]

- The proposed formulation tries to identify relevant dates $d_i$ for $q$;

- **Minimize any errors** that might arise from considering irrelevant or wrong dates.
While for a first order similarity association measure, the definition of a $M_{ct}$ matrix is enough to compute $\text{sim}(q,d_i)$, as the relatedness between the query and the date is simply given by their direct co-occurrence:

<table>
<thead>
<tr>
<th>Haiti Earthquake</th>
<th>1500</th>
<th>1564</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>D</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>
While for a first order sim association measure, the definition of a $M_{ct}$ matrix is enough to compute $\text{sim}(q,d_i)$, as the relatedness between the query and the date is simply given by their direct co-occurrence:

$M_{ct} = [CT_{ij}]_{pxm}$

[www.linkedin.com/in/camposricardo] [www.ccc.ipt.pt/~ricardo]
Temporal Representation of Web Snippets

The use of a **second-order co-occurrence** measure to compute $sim(q,d_i)$ is more complex as it **depends** on setting a **contextual vector for each of the two items** (query and date):

This requires the definition of:

$$M_t = [T_{ij}]_{m \times m}$$

<table>
<thead>
<tr>
<th></th>
<th>1500</th>
<th>1564</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>1500</td>
<td>1</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>1564</td>
<td>v</td>
<td>1</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>2010</td>
<td>v</td>
<td>v</td>
<td>1</td>
<td>v</td>
</tr>
<tr>
<td>2011</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>1</td>
</tr>
</tbody>
</table>
# Conceptual Representation of Web Snippets

$$M_c = [C_{ij}]_{p \times p}$$

<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>
</tr>
</thead>
<tbody>
<tr>
<td>Haiti Earthquake</td>
<td>1</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>major earthquakes</td>
<td>v</td>
<td>1</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>haiti</td>
<td>v</td>
<td>v</td>
<td>1</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>catastrophic damage</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>1</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>port-au-prince</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>1</td>
<td>v</td>
</tr>
<tr>
<td>concepción de la vega</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>1</td>
</tr>
</tbody>
</table>
With regard to the contextual vector, we can define five possible representations:

(q, d) Contextual Vectors Representation

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

With regard to the contextual vector, we can define **five possible representations**:

- \( (D; D) \)

<table>
<thead>
<tr>
<th>Q</th>
<th>( q )</th>
<th>Date</th>
<th>Date</th>
<th>Date</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>( d_i )</td>
<td>Date</td>
<td>Date</td>
<td>Date</td>
<td>…</td>
</tr>
</tbody>
</table>
With regard to the contextual vector, we can define **five possible representations**:

```
(Q;D) Contextual Vectors Representation
```

- **Q**:
  - q
  - Word
  - Word
  - Word
  - ...

- **D**:
  - d_i
  - Date
  - Date
  - Date
  - ...

• (W;D)
With regard to the contextual vector, we can define five possible representations:

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

- \(Q\) → \(q\) Date Date Date …
- \(D\) → \(d_i\) Word Word Word …

\(\cdot (D;W)\)
With regard to the contextual vector, we can define **five possible representations**:

<table>
<thead>
<tr>
<th>Q</th>
<th>→</th>
<th>Word</th>
<th>Date</th>
<th>Word</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>→</td>
<td>Date</td>
<td>Word</td>
<td>Word</td>
<td>…</td>
</tr>
</tbody>
</table>

• \((WD;WD)\)
This work applies the **InfoSimba** (see Dias et. al.) similarity measure, a semantic vector space model supported by corpus-based **word correlations**:

\[
IS(q, d_i) = \frac{\sum_{k \in q} \sum_{j \in d_i} w(k).w(j).S(k,j)}{\left( \sum_{k \in q} \sum_{j \in q} S(k,j) + \right) \left( \sum_{k \in d_i} \sum_{j \in d_i} S(k,j) - \right) \sum_{k \in q} \sum_{j \in d_i} S(k,j)}
\]

In detail, the **InfoSimba** calculates the **correlation** between \( q \) and \( d_i \).

where \( s(k,j) \) is any first order similarity measure (e.g. DICE, PMI) relating words/dates \( k \) and \( j \), and \( w(k) \) and \( w(j) \) correspond to their weights.
Although *IS has shown an improved performance* compared to other state-of-the-art measures *when directly applied to a (q,di) pair*, results were not completely satisfactory.

According to our investigation, it can be concluded, that the relevance between a (q,di) pair is better defined if:

- Instead of just focusing on the self-similarity;
- Compute the similarities between di and W*;

where W* is the set of words W that co-occur with di in any web snippet.
GenTempEval

In order to compute all these values,

<table>
<thead>
<tr>
<th></th>
<th>1500</th>
<th>1564</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haiti Earthquake</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>major earthquakes</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>haiti</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</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>

We propose a new **generic temporal similarity measure** called *GenTempEval*:

\[ GenTempEval(q, d_i) = F(sim(W^*, d_i)) \]

Where **sim** can be any similarity measure, either of **first** or **second-order**;
In order to compute all these values, we propose a new **generic temporal similarity measure** called *GenTempEval*:

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

*F* is an aggregation function that combines the different similarity values \( \text{sim}(W^*, d_i) \) in a single one.
For this purpose, we consider three different $F$ functions:

- Arithmetic Mean;
- Median;
- Max / Min;
42 representative clear-concept implicit temporal queries:

- non-ambiguous in concept;
- temporal in purpose.
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 top best 50 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>
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 \( IS_{(X;Y)}_{SM\_F} \)

- \( IS \to \text{InfoSimba} \);
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;
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;
- $SM \rightarrow$ any similarity measure of first order used with IS (e.g., PMI, DICE,…).
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;
- $SM \rightarrow$ any similarity measure of first order used with IS (e.g., PMI, DICE,…);
- $F \rightarrow$ 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 $IS_{(X;Y)}_{SM}$. 

[www.linkedin.com/in/camposricardo] [www.ccc.ipt.pt/~ricardo]
Baseline Methods

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

Without the aggregator function, denoted SM
Baseline Methods

- PMI_F;
- DICE_F;
- Jaccard_F;
- SCP_F;
- NGoogleDistance_F;
- WebJaccard_F;
- WebOverlap_F;
- WebDICE_F;
- WebPMI_F;

With the aggregator function, denoted SM_F
Comparison of the different Aggregation Functions

For each of the contextual vectors representations, we have performed a set of experiments with different sizes, \( N \)

\[
\begin{array}{c}
q & \ldots & \ldots & \ldots & \ldots \\
d_i & \ldots & \ldots & \ldots & \ldots \\
\end{array}
\]

\( N = ? \)
Comparison of the different Aggregation Functions

For each of the contextual vectors representations, we have performed a set of experiments with different sizes, $N$

<table>
<thead>
<tr>
<th>$q$</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Comparison of the different Aggregation Functions

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

<table>
<thead>
<tr>
<th>Word / Date</th>
<th>( q / d_i )</th>
<th>( q / d_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q / d_i )</td>
<td>( v )</td>
<td>( v &gt; T )</td>
</tr>
</tbody>
</table>

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

\[
\text{Word / Date} \quad q / d_i \quad v
\]

\[
\text{Word / Date} \quad q / d_i \quad v > T
\]

\[
\text{Word / Date} \quad q / d_i \quad v
\]

\[
\text{Word / Date} \quad q / d_i \quad v
\]
Point Biserial Correlation Coefficient

To compare the different measures, we used the point biserial correlation coefficient, a statistical correlation measure that considers items consisting of binary or dichotomous classifications.

<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>

High correlation biserial values indicate high agreement with human annotators.
## Comparison of the different Aggregation Functions

Regardless of the approach used, best correlation coefficient, always occurred with $T_{0.05N^+\infty}$. The best value is given for the Median function, specifically for $IS_{WD;WD}_DICE_M$ denoted BGTE.

<table>
<thead>
<tr>
<th>Aggregation</th>
<th>Measure</th>
<th>N5</th>
<th>N10</th>
<th>N20</th>
<th>$N^+\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max / Min</td>
<td>$IS_{WD;WD}_SCP_M$</td>
<td>0.668</td>
<td>0.708</td>
<td>0.712</td>
<td>0.713</td>
</tr>
<tr>
<td>Mean</td>
<td>$IS_{WD;WD}_DICE_A$</td>
<td>0.550</td>
<td>0.724</td>
<td>0.795</td>
<td>0.799</td>
</tr>
<tr>
<td>Median</td>
<td>$IS_{WD;WD}_DICE_M$</td>
<td>0.540</td>
<td>0.693</td>
<td>0.795</td>
<td><strong>0.800</strong></td>
</tr>
</tbody>
</table>
The type of contextual vector representation chosen for the \((q,d_i)\) pairs greatly influences the performance of the system;

Regardless of the approach used, we found that the best possible representation is given by the combination of words and dates, denoted \((WD;WD)\).

<table>
<thead>
<tr>
<th>Aggregation</th>
<th>((W;W))</th>
<th>((D;D))</th>
<th>((W;D))</th>
<th>((D;W))</th>
<th>((WD;WD))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max / Min</td>
<td>0.706</td>
<td>0.545</td>
<td>0.333</td>
<td>0.449</td>
<td>0.713</td>
</tr>
<tr>
<td>Mean</td>
<td>0.768</td>
<td>0.358</td>
<td>0.387</td>
<td>0.149</td>
<td>0.799</td>
</tr>
<tr>
<td>Median</td>
<td>0.771</td>
<td>0.334</td>
<td>0.366</td>
<td>0.175</td>
<td>0.800</td>
</tr>
</tbody>
</table>
Assess the Temporal Similarity of a \((q,d_i)\) pair

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
  \[ \text{GenTempEval} (q,d_i) \geq \lambda \]
- Irrelevant or wrong date
  \[ \text{GenTempEval} (q,d_i) < \lambda \]
IR Metrics

<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>

We also developed a **ROC curve**. The **red line** indicates an almost perfect **classifier** with an Area Under Curve (AUC) of **0.960**.
We can note a **significant difference** in all the IR measures between using **IS** with the **median paradigm** and **without it** (i.e., direct co-occurrence between \((q,d_i)\):

![Graph showing comparison between IS_(WD;WD)_DICE_M and IS_(WD;WD)_DICE]

**Experiment Setup**

**Conclusions**
Also the Median paradigm reaches high values when compared to the Arithmetic Mean and Max / Min paradigm:

![Graph comparing different aggregation functions](image-url)
When compared to the **baseline measures**, BGTE reaches the highest values:
In our final experiment, we compare the results of the BGTE’s performance with the baseline rule-based model which selects all of the temporal patterns found as correct dates:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Rule-Based Model</th>
<th>BGTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.634</td>
<td>0.748</td>
</tr>
<tr>
<td>Recall</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F1-Measure</td>
<td>0.776</td>
<td>0.856</td>
</tr>
</tbody>
</table>
Furthermore, we assess if the difference between using the BGTE or rule-based model for the correct classification of a pair is significant;

For this purpose we built a confidence interval for the difference of means for paired samples between the number of misclassified dates given by the rule-based method and by the BGTE;

The interval obtained [1.42; 2.30] clearly shows that the rule-based model retrieves, on average, more irrelevant or incorrect dates than the BGTE measure.
We proposed a new temporal similarity measure, the *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 *InfoSimba* with the *DICE* coefficient and the *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.
Future Work

- Use GenTempEval in the field of Temporal Ephemeral Clustering;
- Compare our approach with a query-log based one;
- With a temporal classifier based on multiple similarity measures.
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/