Temporal Information Extraction
Helge Holzmann

Partially based on the tutorial:
‘Temporal Information Extraction and Shallow Temporal Reasoning’
by Dan Roth, Heng Ji, Taylor Cassidy, Quang Do.
24th International Conference on Computational Linguistics (COLING), Mumbai, India, 2012.
1. Motivation and Applications

2. Time in Text / Time Representation

3. Temporal Information from Wikipedia

4. Temporal Slot Filling

5. Evaluation

6. Demo / Experiments
1. Motivation and Applications

2. Time in Text / Time Representation

3. Temporal Information from Wikipedia

4. Temporal Slot Filling

5. Evaluation

6. Demo / Experiments
(Temporal) Information Extraction

- Information Extraction
  - Extraction of structured information from unstructured text
  - Relations, events, facts (e.g., <subject> <predicate> <object>)
  - Many relations and events are temporally bounded
    - a person's place of residence or employer
    - an organization's members
    - the duration of a war between two countries
    - the precise time at which a plane landed
    - ...

- Temporal Information Extraction
  - Temporal Expressions
    - Dates
    - Times
    - Relative expressions (e.g., yesterday)
  - Temporal Information
    - Information with time attached
    - Temporal Slot Filling (TSF)
      - Connecting existing information with time
Motivation: TIE in Temporal IR

- Applications
  - Knowledge base population
    - Temporal knowledge bases
  - Disambiguation / Linking
    - Time-aware linking / disambiguation
  - Entity graphs
    - Evolving entity graphs
  - Entity based indexing
    - Temporal indexes / IR
  - Temporal query suggestions / diversification
    - Time travel queries
  - ...

Helge Holzmann 05/01/15
Introducing the Knowledge Graph: things, not strings

Cross-posted on the Inside Search Blog

Search is a lot about discovery—the basic human need to learn and broaden your horizons. But searching still requires a lot of hard work by you, the user. So today I’m really excited to launch the Knowledge Graph, which will help you discover new information quickly and easily.

Take a query like [taj mahal]. For more than four decades, search has essentially been about matching keywords to queries. To a search engine the words [taj mahal] have been just that—two words.

But we all know that [taj mahal] has a much richer meaning. You might think of one of the world’s most beautiful monuments, or a Grammy Award-winning musician, or possibly even a casino in Atlantic City, NJ. Or, depending on when you last ate, the nearest Indian restaurant. It’s why we’ve been working on an intelligent model—in geek-speak, a “graph”—that understands real-world entities and their relationships to one another: things, not strings.

The Knowledge Graph enables you to search for things, people or places that Google knows about—landmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more—and instantly get information that’s relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do.
Motivation: TIE in IR
German reunification (Deutsche Wiedervereinigung) took place on **October 3, 1990**, when the areas of the former German Democratic Republic (GDR, in English commonly called "East Germany") were incorporated into the Federal Republic of Germany (FRG, in English commonly called "West Germany"), both formed in 1949, after ...

**German reunification - New World Encyclopedia**
www.newworldencyclopedia.org/entry/German_reunification

**German reunification - Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/German_reunification

The consensus opinion was that reunification, if it must occur, should not occur until at least 1995 .... However, German separation did not result in another war. City Palace, Berlin - Removal of Hungary's border ... - Palace of the Republic (Berlin)

**German reunification - New World Encyclopedia**
www.newworldencyclopedia.org/entry/German_reunification

Doc 11, 2013 - German reunification (Deutsche Wiedervereinigung) took place on **October 3, 1990**, when the areas of the former German Democratic Republic (GDR, in English commonly called "East Germany") were incorporated into the Federal Republic of Germany (FRG, in English commonly called "West Germany"), both formed in 1949, after ...

**Reunification of Germany 3 October 1990 | ICV**
vc.vlada.cz/en/tema/reunification-of-germany-3-october.../tmpid=576/

Oct 3, 1990 - The reunion of Germany in 1990 is one of the most important historical ... At that time Kohl did not believe that he would ever negotiate with Mikhall ... Further important shift in opinion took place during the visit of Helmut ...
Obama was re-elected president in November 2012, defeating Republican nominee Mitt Romney, and was sworn in for a second term on January 20, 2013.

Reference: en.wikipedia.org/wiki/President_Barak_Obama

Barack Obama - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/President_Barak_Obama
The inauguration of Barack Obama as the 44th President took place on ... Obama became the first President of the United States to address both houses of the UK ... Early life and career · Legislative career ... · Presidential campaigns

President Barack Obama | The White House
www.whitehouse.gov/administration/president-obama
Official biography of President Barack Obama

How Barack Obama Became the Anti-President - US News
www.usnews.com/.../12/05/how-barack-obama-became-the-anti-president
Dec 05, 2014 · President Barack Obama's aloof style is responsible for the cruel slide from hope and change to partisan gridlock.

Barack Obama elected 44th president - NBC News - ...
www.nbcnews.com/.../barack-obama-elected-th-president
Barack Obama shattered more than 200 years of history ... President-elect Barack Obama walks on stage at his ... was seeking to become the nation's ...
Knowledge Bases
Arnold Alois Schwarzenegger (German: Arnold Alois Schwarzenegger; born July 30, 1947) is an Austrian-born American actor, film producer, businessman, investor, writer, philanthropist, former professional bodybuilder and politician. Schwarzenegger served two terms as the 38th Governor of California from 2003 until 2011. Schwarzenegger began weight training at the age of 16. He won the Mr. Universe title at age 20 and went on to win the Mr. Olympia contest seven times.

Property | Value
--- | ---
dbpedia-owl:abstract | Arnold Alois Schwarzenegger (German: Arnold Alois Schwarzenegger; born July 30, 1947) is an Austrian-born American actor, film producer, businessman, investor, writer, philanthropist, former professional bodybuilder and politician. Schwarzenegger served two terms as the 38th Governor of California from 2003 until 2011. Schwarzenegger began weight training at the age of 16. He won the Mr. Universe title at age 20 and went on to win the Mr. Olympia contest seven times. Schwarzenegger has remained a prominent presence in bodybuilding and has written many books and articles on the sport. Schwarzenegger gained worldwide fame as a Hollywood action film icon. He was nicknamed the "Austrian Oak" and the "Styrian Oak" in his bodybuilding days, "Arnie" during his acting career and more recently "The Governor" (a portmanteau of "Governor" and "The Terminator" – one of his best-known movie roles). As a Republican, he was first elected on October 7, 2003, in a special recall election to replace then-Governor Gray Davis. Schwarzenegger was sworn in on November 17, 2003, to serve the remainder of Davis's term. Schwarzenegger was then re-elected on November 7, 2006, in California’s 2006 gubernatorial election, to serve a full term as governor, defeating Democrat Phil Angelides, who was California State Treasurer at the time. Schwarzenegger was sworn in for his second term on January 6, 2007. In 2011, Schwarzenegger completed his second term as governor, and it was announced that he had separated from Maria Shriver, his wife for the last 25 years; she is a member of the influential Kennedy family, as a niece of the late Democratic US President John F. Kennedy.

dbpedia-owl:activeYearsEndDate | 2011-01-03 (xsd:date)
dbpedia-owl:activeYearsStartDate | 2003-11-17 (xsd:date)
dbpedia-owl:alias | Schwarzenegger, Arnold Alois
dbpedia-owl:birthDate | 1947-07-03 (xsd:date)
dbpedia-owl:birthName | Arnold Alois Schwarzenegger
dbpedia-owl:birthPlace | dbpedia:Austria
dbpedia-owl:birthYear | 1947-01-01 (xsd:date)
dbpedia-owl:child | dbpedia:Katharine_Schwarzenegger
dbpedia-owl:lieutenant | dbpedia:Abel_Maldonado
dbpedia-owl:lieutenant | dbpedia:Cruz_Rustamante
dbpedia-owl:lieutenant | dbpedia:John_Garamendi
dbpedia-owl:lieutenant | dbpedia:Monica_Pasquel
dbpedia-owl:militaryBranch | dbpedia:Austrian_Armed_Force
dbpedia-owl:occupation | dbpedia:Bodybuilder
dbpedia-owl:office | dbpedia:Governor_of_California
dbpedia-owl:orderInOffice | 38th
dbpedia-owl:party | dbpedia:Republican_Party_(United_States)
dbpedia-owl:profession | dbpedia:Bodybuilder
dbpedia-owl:relation | dbpedia:Gustav_Schwarzenegger
dbpedia-owl:religion | dbpedia:Catholic_Catholic
dbpedia-owl:residence | dbpedia:California
dbpedia-owl:residence | dbpedia:Brentwood,_Los_Angeles
dbpedia-owl:serviceEndDate | 1965-01-01 (xsd:date)
dbpedia-owl:serviceStartDate | 1965-01-01 (xsd:date)
dbpedia-owl:successor | dbpedia:Jerry_Brown
dbpedia-owl:termEndDate | dbpedia:Arnold_Schwarzenegger_1
dbpedia-owl:termStartDate | dbpedia:Arnold_Schwarzenegger_2
dbpedia-owl:wikiPageExternalLink | http://gov36.ca.gov/
dbpedia-owl:wikiPageExternalLink | http://www.life.com/gallery/60601/arnold-schwarzenegger-wild-years#index/0
(Temporal) Knowledge
Renamings and Temporal Translation

Example: Exxon Valdez Oil Spill 1989

- Right after the spill Exxon Valdez was renamed:
  - Exxon Mediterranean
- Ever since, 5 more times:
  - Sea River Mediterranean
  - S/R Mediterranean
  - Mediterranean
  - Dong Fang Ocean
  - Oriental Nicety

You want to search for Oriental Nicety today
1. Motivation and Applications

2. Time in Text / Time Representation

3. Temporal Information from Wikipedia

4. Temporal Slot Filling

5. Evaluation

6. Demo / Experiments
**Time in Text**

- **Temporal Expressions**
  - in 1992; 1990’s; on Dec. 8th, 1992; at 8:00 am
  - from Monday to Friday; between 9 and 10 am
  - this/next Century; year; day; hour; minute; second, …
  - before 1992; until next year; for (about) one year;
  - tomorrow; after 8:00;
  - Tuesdays; first of the month; several times per week;

- **Denote intervals and points at varying granularity and specificity**
  - Now (December 8th, 2012, 3:00pm)
    - 2012-12-08-T15:00
  - December 8th, 2012
    - 2012-12-08-TXX:XX
    - (2012-12-08T00:00, 2012-12-08-T11:59)
  - December 8th
    - xxxx-12-08-Txx:xx
  - Set of all December 8th’s
  - Each is an interval: { (xxxx-12-08:T00:00, xxxx-12-08:T11:59) }
Temporal Expressions

- Reference Date: January 5, 2015
  - E.g., document creation date, publication date, upload date, …

<table>
<thead>
<tr>
<th>Expression</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 5, 2015</td>
<td>2015-01-05</td>
</tr>
<tr>
<td>Friday</td>
<td>2015-01-02</td>
</tr>
<tr>
<td>today</td>
<td>2015-01-05</td>
</tr>
<tr>
<td>1993</td>
<td>1993</td>
</tr>
<tr>
<td>the 1990’s</td>
<td>199X</td>
</tr>
<tr>
<td>midnight, December 8, 2012</td>
<td>2012-12-08T00:00:00</td>
</tr>
<tr>
<td>5pm</td>
<td>2015-01-05T17:00</td>
</tr>
<tr>
<td>the previous day</td>
<td>2015-01-04</td>
</tr>
<tr>
<td>last October</td>
<td>2014-10</td>
</tr>
<tr>
<td>last autumn</td>
<td>2014-FA</td>
</tr>
<tr>
<td>last week</td>
<td>2014-W52</td>
</tr>
<tr>
<td>Thursday evening</td>
<td>2015-01-01TEV</td>
</tr>
<tr>
<td>two months ago</td>
<td>2014-11</td>
</tr>
</tbody>
</table>
Temporal Expression Extraction

- Rule-based [Strötgen and Gertz, 2010; Chang and Manning, 2012; Do et al., 2012]
- Machine Learning
  - Risk Minimization Model [Boguraev and Ando, 2005]
  - Conditional Random Fields [Ahn et al., 2005; UzZaman and Allen, 2010]

- HeidelTime [Strötgen and Gertz, 2010]
  - Three ways to describe temporal expressions [Schilder and Habel, 2001]
    - explicitly
      - e.g., January 5, 2015 $\rightarrow$ 2015-01-05
    - implicitly
      - e.g., Christmas Eve 2014 $\rightarrow$ 2014-12-24
    - relatively
      - e.g., last June $\rightarrow$ (YYYY-06-01, YYYY-06-30)
      - YYYY depends on the reference time, either:
        - document creation time (dct)
        - previously mentioned date
        - ...

Eventualities

- Temporal expressions describe temporal information / eventualities
  - events, processes, states, happenings, changes, episodes, etc.
- Types of eventualities [Dölling, 2011]
Relations of Eventualities

- Types of eventualities [Dölling, 2011]
  - A boundary begins/ends a happening
  - A boundary culminates an event
  - A moment is the reduction of an episode
  - A state is the result of a change
  - A habitual state is realized by a class of occurrences
  - A process is made of event constituents

- Task: connect eventualities in text with temporal expressions
1. Motivation and Applications

2. Time in Text / Time Representation

3. Temporal Information from Wikipedia

4. Temporal Slot Filling

5. Evaluation

6. Demo / Experiments
Temporal Information from Wikipedia

- **YAGO2**: A spatially and temporally enhanced knowledge base from Wikipedia [Hoffart et al., 2013]
  - Temporal information for entities and facts
    - Entities: existence
      - *people, groups, artifacts, events*
    - Facts: time points for events, time spans for validity
      - e.g., *was born in, holds political position* (e.g., president)
  - Resolution: one day or less
    - Format: *YYYY-MM-DD, YYYY-##-##*
  - Wikipedia as source
    - Infoboxes: e.g., birth date
    - Categories: e.g., ‘82nd Academy Awards’ in ‘2009 Film Awards’
  - Already known dates of subject or object of a fact: e.g.,
    - *BobDylan wasBornIn Duluth*, birth date of Bob Dylan
    - *ElvisPresley diedIn Memphis*, death date of Elvis Presley
    - *BobDylan created BlondeOnBlonde*, creation date of ‘Blonde On Blonde’

YAGO, [http://www.yago-knowledge.org](http://www.yago-knowledge.org)
Entities in YAGO2

- Three kinds of existence relations
  - `startsExistingOnDate`
  - `endsExistingOnDate`
  - `happenedOnDate`

- Existence of different entity types
  - **People**
    - from *birth* to *death*
  - **Groups** (music band, football clubs, universities, companies, etc.)
    - from *creation* to *destruction*
  - **Artifacts** (buildings, paintings, books, music songs, etc.)
    - from *creation* to *destruction*
  - **Events** (wars, sports competitions, named epochs, etc.)
    - from *started* to *ended*
    - *happened* if started and ended at the same day

---

**82nd Academy Awards**

From Wikipedia, the free encyclopedia

| Categories: Academy Awards ceremonies | 2009 film awards | 2010 in California | 21st century in Los Angeles, Calif |

---

**Steve Jobs**

Jobs holding an iPhone 4 at Worldwide Developers Conference 2010

- **Born**
  - Steven Paul Jobs
  - February 24, 1955
  - San Francisco, California, US
- **Died**
  - October 5, 2011 (aged 56)
  - Palo Alto, California, US
- **Residence**
  - Palo Alto, California, US
Facts in YAGO2

- Three kinds of occurrence relations
  - `occursSince`
  - `occursUntil`
  - `occursOnDate`

- Facts are assigned #IDs and used as subject
  - #1: BarackObama wasInauguratedAs PresidentOfTheUnitedStates
    - #1 occursOnDate 2009-01-20
      - #1 occursSince 2009-01-20
      - #1 occursUntil 2009-01-20
  - #2: BobDylan hasWonPrize GrammyAward
    - #2 occursOnDate 1973-##-##
  - #3: BobDylan hasWonPrize GrammyAward
    - #3 occursOnDate 1979-##-##
  - #4: BradPitt spouse JenniferAniston
    - #4 occursSince 2000-##-##
    - #4 occursUntil 2005-##-##
  - #5: BradPitt spouse AngelinaJolie
    - #5 occursSince 2014-##-##
Facts with a deduced time in YAGO2

- Temporal information are deduced from existence of subject or object of a fact
- Based on rules
  - Permanent relations
    - identifying attributes, like ID numbers or types (e.g., `sb isA person`)
      
      $$
      \begin{align*}
      \text{id} & : \text{s} \ \text{p} \ \text{o}; \\
      \text{p} & \text{ type permanentRelation;}
      \text{s} & \text{ startsExistingOnDate} \ \text{b} ;
      \text{e} & \text{ endsExistingOnDate} \\
      \text{id} & \text{ occursSince} \ \text{b}; \\
      \text{id} & \text{ occursUntil} \ \text{e}
      \end{align*}
      $$
  
  - Creation relations
    - e.g., `ElvisPresley wasBornIn Tupelo`, birth date of Elvis Presley (subject)
    - e.g., `LeonardCohen created Suzanne(song)`, creation date of Suzanne
      
      $$
      \begin{align*}
      \text{id} & : \text{s} \ \text{p} \ \text{o}; \\
      \text{p} & \text{ type objectStartRelation;}
      \text{o} & \text{ startsExistingOn} \ \text{b} \\
      \text{id} & \text{ occursSince} \ \text{b}; \\
      \text{id} & \text{ occursUntil} \ \text{b}
      \end{align*}
      $$
  
  - Destruction relations
    - e.g., `Taliban destroyed BuddhasOfBamyan`, destruction date of Buddhas
      
      $$
      \begin{align*}
      \text{id} & : \text{s} \ \text{p} \ \text{o}; \\
      \text{p} & \text{ type subjectEndRelation;}
      \text{s} & \text{ endsExistingOn} \ \text{e} \\
      \text{id} & \text{ occursSince} \ \text{e}; \\
      \text{id} & \text{ occursUntil} \ \text{e}
      \end{align*}
      $$
1. Motivation and Applications

2. Time in Text / Time Representation

3. Temporal Information from Wikipedia

4. Temporal Slot Filling

5. Evaluation

6. Demo / Experiments
**Incomplete / Fuzzy Times**

**Challenge**
- Times are incomplete
- Times are mentioned with different granularities
- Uncertainty in time information

**Solution: 4-tuple representation [Ji et al., 2011]**
- \(<t_1, t_2, t_3, t_4>\)
  - \(t_1 < t_{\text{start}} < t_2\)
  - \(t_3 < t_{\text{end}} < t_4\)

<table>
<thead>
<tr>
<th>Document text (2001-01-01)</th>
<th>(t_1)</th>
<th>(t_2)</th>
<th>(t_3)</th>
<th>(t_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chairman Smith</td>
<td>-infinite</td>
<td>2001-01-01</td>
<td>2001-01-01</td>
<td>+infinite</td>
</tr>
<tr>
<td>Smith, who has been chairman for two years</td>
<td>-infinite</td>
<td>1999-01-01</td>
<td>2001-01-01</td>
<td>+infinite</td>
</tr>
<tr>
<td>Smith, who was named chairman two years ago</td>
<td>1999-01-01</td>
<td>1999-01-01</td>
<td>1999-01-01</td>
<td>+infinite</td>
</tr>
<tr>
<td>Smith, who resigned last October</td>
<td>-infinite</td>
<td>2000-10-01</td>
<td>2000-10-01</td>
<td>2000-10-31</td>
</tr>
<tr>
<td>Smith served as chairman for 7 years before leaving in 1991</td>
<td>1984-01-01</td>
<td>1984-12-31</td>
<td>1991-01-01</td>
<td>1991-12-31</td>
</tr>
<tr>
<td>Smith was named chairman in 1980</td>
<td>1980-01-01</td>
<td>1980-12-31</td>
<td>1980-01-01</td>
<td>+infinite</td>
</tr>
</tbody>
</table>
Temporal Slot Filling (TSF)

“Given a query entity, a knowledge base (KB) and a source corpus, a system must return slot fills and temporal information must be gathered across the entire corpus” [Ji et al., 2013]

Example:

- <Google> <CEO> <Eric Schmidt>
  - “Schmidt joined Google’s board of directors as chairman in March 2001, and became the company’s CEO in August 2001.” (Wikipedia)
    - Start: August 2001 (2001-08-01, 2001-08-31)
  - “Schmidt is one of a few people who became billionaires (in United States dollars) based on stock options received as employees in corporations of which they were neither the founders nor relatives of the founders. In its 2011 ‘World’s Billionaires’ list, Forbes ranked Schmidt as the 136th-richest person in the world, with an estimated wealth of $7 billion. Google gave him a $100 million equity award in 2011 when he stepped down as CEO.” (Wikipedia)
  - 4-tuple: <2001-08-01, 2001-08-31, 2011-01-01, 2011-12-31>
  - Two steps: Temporal Classification, Temporal Aggregation
Temporal Classification

- Identify text instances containing
  - Information, e.g., fact: <subject> <predicate> <object>
  - Temporal expression

- Classify as
  - START, END, HOLDS, RANGE, NONE

Examples
- START
  - Rob joined Microsoft in 1999.
- END
  - Rob left Microsoft in 1999.
- HOLDS
  - In 1999 Rob was still working for Microsoft.
- RANGE
  - Rob has worked for Microsoft for the last ten years.
- NONE
  - Last Sunday Rob’s friend joined Microsoft.
Temporal Aggregation

- **Normalization**
  - E.g., ‘he got married in 2014’
    - certain: marriage in (2014-01-01, 2014-12-31) = (t_a, t_b)
  - Uncertainty
    - divorce already at t_a? Unlikely, but we do not know
    - maybe still married → +INF

- **Creating 4-tuple corresponding to the identified classes**
  - **START:** <t_a, t_b, t_a, +INF>
  - **END:** <-INF, t_b, t_a, t_b>
  - **HOLDS:** <-INF, t_a, t_b, +INF>
  - **RANGE:** <t_a, t_b, t_a, t_b>
  - **NONE:** <-INF, +INF, -INF, +INF>

- **Aggregation by combining every two pairs of tuples T and T’**
- **Iterative aggregation [Li et al., 2012]**

\[ T \land T' = < \max(t_1, t'_1), \min(t_2, t'_2), \max(t_3, t'_3), \min(t_4, t'_4) > \]
Flat Approach

- Extract temporal expressions, normalize
  - ‘In 1981, Makoni was moved to the position of minister of industry and energy development, where he remained until 1983.’
    - Entity: Makoni
    - Attribute / value: position / minister of industry and energy development
    - Target temporal expression: 1983
  - ‘In DATE, TE was moved to the position of TA, where he remained until TD.’
    - Extract window
    - Incorporate dependency features (dependency parsing)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE Win</td>
<td>be, move, to, in, DATE, the</td>
</tr>
<tr>
<td>TA Win</td>
<td>of, to, remain, position, the, where, until, he</td>
</tr>
<tr>
<td>TD Win</td>
<td>remain, where, until, he</td>
</tr>
<tr>
<td>TE Governs</td>
<td>-</td>
</tr>
<tr>
<td>TA Governs</td>
<td>-</td>
</tr>
<tr>
<td>TD Governs</td>
<td>-</td>
</tr>
<tr>
<td>TE Governed by</td>
<td>move</td>
</tr>
<tr>
<td>TA Governed by</td>
<td>position</td>
</tr>
<tr>
<td>TD Governed by</td>
<td>remain</td>
</tr>
</tbody>
</table>

[Li et al., 2012]
Long Contexts

- Biggest TSF challenge
  - Sentences cover more than one information / fact
  - Facts span more than one sentence
  - Temporal information are not always close to the facts

‘In August 2000, as a result of the conclusion of arbitration with the International Chamber of Commerce, Andersen Consulting broke all contractual ties with AWSC and Arthur Andersen. As part of the arbitration settlement, Andersen Consulting paid the sum held in escrow (then $1.2 billion) to Arthur Andersen, and was required to change its name, resulting in the entity being renamed Accenture.’

- Solution
  - Use parsing structures to compress long contexts
    - Dependency parsing / co-reference resolution
      - Far from perfect, not robust enough
    - Combine structured and flat approaches [Li et al., 2012]
Structured Approach

- Representation based on three shortest dependency paths

‘In 1975, after being fired from Columbia amid allegations that he used company funds to pay for his son's bar mitzvah, Davis founded Arista.’

- Surface sentence
  - Long distance between 1975 and Davis founded Arista
  - Some words in between cause ambiguity: fired

- Dependency paths
  - Help remove irrelevant information
  - Build syntactic and semantic links from long distance
    - Requires text analysis / Natural Language Processing (NLP)

[Li et al., 2012]
‘In September 2005, Dichter left office and became a research fellow at the Brookings Institute in Washington, D.C.’

Part-of-Speech:

```
1  In September 2005, Dichter left office and became a research fellow at the Brookings Institute in Washington, D.C.
```

Named Entity Recognition:

```
1  In September 2005, Dichter left office and became a research fellow at the Brookings Institute in Washington, D.C.
```
‘In September 2005, Dichter left office and became a research fellow at the Brookings Institute in Washington, D.C.’

Basic dependencies:

- No connection between September 2005 and Brookings Institute Washington
  - Flat approach works, but structured does not
  - However, with long contexts, structured often works better
    - Combine flat and structured approach [Li et al., 2012]

Stanford CoreNLP, http://nlp.stanford.edu:8080/corenlp/process
Training Set Creation

- Manual annotation
  - Requires experts
  - Much effort, very time-consuming task
  - Explicit information is very sparse
    - Scattered across documents

- Alternative: Distant Supervision [Mintz et al., 2009]
  - Retrieve temporal information from a Knowledge Base (KB)
  - Collect top web search results for ‘John Mary’
    - www.mary-and-john.com
    - www.mymarriage.com/john33/
    - www.wedding-photos.com/joma/
    - ...
  - Collect all sentences that mention both entities
    - ‘On 1997, John and Mary renewed their vows in Florida.’
  - Compare with KB, label expressions
    - Sentence: 1997 → START
Tim Cook is the CEO of Apple and serves on its Board of Directors. Before being named CEO in August 2011, Tim was Apple's Chief Operating Officer and was responsible ...
1. Motivation and Applications

2. Time in Text / Time Representation

3. Temporal Information from Wikipedia

4. Temporal Slot Filling

5. Evaluation

6. Demo / Experiments
### Evaluation

- **Temporal Quality**
  - Let \(<t_1, t_2, t_3, t_4>\) be system output, \(<g_1, g_2, g_3, g_4>\) be gold standard

  \[
  Q(S) = \frac{1}{4} \sum_{i} c \left( \frac{c + |t_i - g_i|}{c} \right)
  \]

  - An error of \(c\) time units produces a 0.5 score
  - Each element in tuple is scored independently
  - A correct slot fill with temporal information \(t\) gets credit \(Q(S)\) (instead of 1)

- **Overall Metrics (Precision / Recall)**

  \[
  P = \frac{\sum_{S^i \in C(S)} Q(S^i)}{M} \quad R = \frac{\sum_{S^i \in C(S)} Q(S^i)}{N}
  \]

  - \(M\): the number of system output tuples
  - \(N\): the number of gold standard tuples
  - \(C(S)\): the number of instances that have correct slot fills
A parameter should determine if a certain amount of vagueness is worse/better than a certain amount of over-constraining

\[ c_i = \begin{cases} 
  c_{vag}, & \text{if } (i \in \{1,3\} \land t_i \leq g_i) \lor (i \in \{2,4\} \land t_i \geq g_i) \\
  c_{cons}, & \text{otherwise}
\end{cases} \]

If \( S \subset S_g \):

\[ Q(S) = \frac{1}{4} \sum_i \frac{c_{cons}}{c_{vag} + |t_i - g_i|} \]

If \( S_g \subset S' \):

\[ Q(S') = \frac{1}{4} \sum_i \frac{c_{vag}}{c_{vag} + |t_i - g_i|} \]
Evaluation Example

- $c_{\text{cons}} = c_{\text{vag}} = 5$
- $+\text{Infinite} = 10,000$
- $-\text{Infinite} = 0$
1. Motivation and Applications
2. Time in Text / Time Representation
3. Temporal Information from Wikipedia
4. Temporal Slot Filling
5. Evaluation
6. Demo / Experiments
Demo / Experiments

- TSF for ‘Places lived’ of Elvis Presley
  - http://www.freebase.com/m/02jq1
Demo / Experiments

- Distant Supervision using William Shakespeare
  - [http://www.freebase.com/m/081k8](http://www.freebase.com/m/081k8)
Distant Supervision using William Shakespeare

“... Stratford-upon-Avon, Warwickshire, England, where it is believed that William Shakespeare was born in 1564...” ➔ START

January

January 1: It is 30 years since the first mobile phone call in the UK was made by comedian Ernie Wise.

He called Vodafone’s head office in Berkshire from St Katherine Docks in London.

January 8: 80 years ago Elvis Presley, the “King” of rock and roll, was born in Tupelo, Mississippi.

His identical twin brother Jesse was delivered stillborn 35 minutes before him.

January 10: The Sinclair C5, a battery-powered electric vehicle, was launched 30 years ago to immediate ridicule – but it did help earn its inventor Sir Clive Sinclair his knighthood.

January 24: The 50th anniversary of the death of Sir Winston Churchill who died at the age of 90 at his London home after suffering a stroke.

January 27: 70 years since the Soviet Red Army liberated Auschwitz concentration camp in southern Poland where at least 1.1 million prisoners had died.

February

Why 2015 is a year to remember: A guide to significant anniversaries

FROM the birth of Elvis Presley to the death of Sir Winston Churchill and the first episode of EastEnders, here’s our guide to the anniversaries coming up over the next 12 months.

BY JANE WARREN

Reference date?

- Document creation date: 2014-12-27
- Actual reference: 2015-01-08
Demo / Experiments

- [http://heideltime.ifi.uni-heidelberg.de/heideltime/](http://heideltime.ifi.uni-heidelberg.de/heideltime/)

HeidelTime Demo

**Configuration**

- HeidelTime is a multilingual and cross-domain temporal tagger.
- Please click the question marks for additional instructions.
- Document type: News
- Language: English
- Document creation time: 2015-01-01

**Input**

- Choose between manually entering text and inserting a text file (up to 2 MB, ensure it's encoded in UTF-8).

**Output**

- Extracted temporal expressions are marked in blue. To see their normalization value, click them.
- You may also receive a TimeML-annotated file (ensure your browser isn't blocking popups).

- Yes, I want to receive a TimeML-annotated file.

- Type: DATE
  - Value: 1935
  - 82 years ago: Elvis Presley, the "King" of rock and roll, was born in Tupelo, Mississippi.
Demo / Experiments


**Stanford CoreNLP**

Output format: Visualise ▼

Please enter your text here:

80 years ago Elvis Presley, the "King" of rock and roll, was born in Tupelo, Mississippi.

Submit ▼ Clear

**Part-of-Speech:**

1. 80 years ago Elvis Presley, the 'King' of rock and roll, was born in Tupelo, Mississippi.

**Named Entity Recognition:**

1. 80 years ago Elvis Presley, the 'King' of rock and roll, was born in Tupelo, Mississippi.

**Coreference:**

1. 80 years ago Elvis Presley, the 'King' of rock and roll, was born in Tupelo, Mississippi.

**Basic dependencies:**

1. 80 years ago Elvis Presley, the 'King' of rock and roll, was born in Tupelo, Mississippi.
### Alternative: use rules (compare YAGO2)


**About: Elvis Presley**

An Entry of Type: [instance person](http://dbpedia.org) within Data Space: [dbpedia.org]

Elvis Aaron Presley (January 8, 1935 – August 16, 1977) was an American singer, musician, and actor. Regarded as one of the most significant cultural icons of the 20th century, he is often referred to as the “King of Rock and Roll” or simply, the “King” born in Tupelo, Mississippi. Presley and his family moved to Memphis, Tennessee, when he was 13 years old. His music career began there in 1954, when he started to work with Sam Phillips, the owner of Sun Records.

- [Elvis Presley](http://dbpedia.org/page/Elvis_Presley)

#### Property-Value Table

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>birthPlace</td>
<td>Tupelo, Mississippi, United States</td>
</tr>
<tr>
<td>branch</td>
<td>United States Army</td>
</tr>
</tbody>
</table>

**TSF value equals birth place (creation relation)**

- Elvis Aaron Presley
- Tupelo, Mississippi, United States
- United States Army
Elvis Presley existence

<table>
<thead>
<tr>
<th>dbpedia-owl:exists</th>
<th>Elvis Presley</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl:birthDate</td>
<td>1935-01-08 (xsd date)</td>
</tr>
<tr>
<td>dbpedia-owl:birthName</td>
<td>Elvis Aaron Presley</td>
</tr>
<tr>
<td>dbpedia-owl:birthPlace</td>
<td>Tupelo, Mississippi</td>
</tr>
<tr>
<td>dbpedia-owl:birthYear</td>
<td>1935-01-01 (xsd date)</td>
</tr>
<tr>
<td>dbpedia-owl:child</td>
<td>Lisa Marie Presley</td>
</tr>
<tr>
<td>dbpedia-owl:deathDate</td>
<td>1977-08-16 (xsd date)</td>
</tr>
<tr>
<td>dbpedia-owl:deathPlace</td>
<td>Memphis, Tennessee</td>
</tr>
<tr>
<td>dbpedia-owl:deathYear</td>
<td>1977-01-01 (xsd date)</td>
</tr>
<tr>
<td>dbpedia-owl:gender</td>
<td>Male</td>
</tr>
</tbody>
</table>

- \[\text{startsExistingOn} \ 1935-01-08\]
- \[\text{endsExistingOn} \ 1977-08-16\]

Creation relation rule

\[
\begin{align*}
\text{id} & : \text{s} \ p \ \text{so} ; \\
\text{p \ type \ objectStartRelation} ; \\
\text{so \ startsExistingOn} \ \text{sb} \\
\text{id \ occursSince} \ \text{sb} ; \\
\text{id \ occursUntil} \ \text{sb}
\end{align*}
\]

- \#1: ElvisPresley birthPlace Tupelo
  - \#1 occursSince 1935-01-08
  - \#1 occursUntil 1935-01-08

\rightarrow \ #2: ElvisPresley placesLived Tupelo
  - \#2 occursSince 1935-01-08
Demo / Experiments

- Evaluation *(Resolution: month)*
  - \( c = 1, +\text{INF} = 100 \)
  
  \[
  Q(S) = \frac{1}{4} \sum_{i} \left( \frac{c}{c+|t_i-g_i|} \right) 
  \]

  \[
  P = \frac{\sum_{S^i \in C(S)} Q(S^i)}{M} 
  \]

  \[
  R = \frac{\sum_{S^i \in C(S)} Q(S^i)}{N} 
  \]

- \( g = \langle 1935-01-08, 1935-01-08, 1935-01-08, +\text{INF} \rangle \)

- \( t = \langle 1935-01-01, 1935-12-31, 1935-01-01, +\text{INF} \rangle \)

- \( Q(S) = \frac{1}{4} \cdot \left( \frac{1}{1+|1-1|} + \frac{1}{1+|12-1|} + \frac{1}{1+|1-1|} + \frac{1}{1+|100-100|} \right) = \frac{1}{4} \cdot \left( 1 + \frac{1}{12} + 1 + 1 \right) = \frac{37}{48} = 0.77 \)

- \( P = R = Q(S) \) *(because M = N = 1)*
References

[Ahm et al., 2005]

[Boguraev and Ando, 2005]

[Chang and Manning, 2012]

[Do et al., 2012]
 References

- [Dölling, 2011]

- [Hoffart et al., 2013]

- [Ji et al., 2011]

- [Ji et al., 2013]

- [Li et al., 2012]
References

- [Mintz et al., 2009]

- [Schilder and Habel, 2001]

- [Strötgen and Gertz, 2010]

- [UzZaman and Allen, 2010]