Information extraction for temporal question answering

David Ahn

ISLA
Universiteit van Amsterdam

SIKS-course Information Retrieval
May 24, 2007

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Kinds of temporal questions

Questions that ask for times

- When did J.R.R. Tolkein die?

Questions that ask for time-dependent information

- Who was the president of the U.S. in 1958?

Questions that ask for event-dependent information

- Which U.S. President resigned in the wake of the Watergate scandal?

Questions that ask about generic temporal sequences

- Is it better to eat before or after a workout?

Kinds of temporal questions

Questions that ask for times

- When did J.R.R. Tolkein die?

Questions that ask for time-dependent information

- Who was the president of the U.S. in 1958?

Questions that ask for event-dependent information

- Which U.S. President resigned in the wake of the Watergate scandal?

Questions that ask about generic temporal sequences

- Is it better to eat before or after a workout?

Kinds of temporal questions

Questions that ask for times

- When did J.R.R. Tolkein die?

Questions that ask for time-dependent information

- Who was the president of the U.S. in 1958?

Questions that ask for event-dependent information

- Which U.S. President resigned in the wake of the Watergate scandal?

Questions that ask about generic temporal sequences

- Is it better to eat before or after a workout?

Kinds of temporal questions

Questions that ask for times

- When did J.R.R. Tolkein die?

Questions that ask for time-dependent information

- Who was the president of the U.S. in 1958?

Questions that ask for event-dependent information

- Which U.S. President resigned in the wake of the Watergate scandal?

Questions that ask about generic temporal sequences

- Is it better to eat before or after a workout?

Kinds of temporal questions

Questions that ask for times

- When did J.R.R. Tolkein die?

Questions that ask for time-dependent information

- Who was the president of the U.S. in 1958?

Questions that ask for event-dependent information

- Which U.S. President resigned in the wake of the Watergate scandal?

Questions that ask about generic temporal sequences

- Is it better to eat before or after a workout?

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
  - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Our focus: temporally restricted questions (time- and event-dependent questions)

Processes necessary for answering temporally restricted questions:

- Identification of dates and times
- Identification of events and states
    - Including information about participants, etc.
- Identification of temporal relations
- Reasoning about temporal relations

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* $\rightarrow$ after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp Pim Fortuyn killing, President Nixon resignation

Vague Watergate scandal

Temporal relations can be hard to identify

Specific *before*, *after*

Vague *in the wake of* $\rightarrow$ after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp	Pim Fortuyn killing, President Nixon resignation
Vague	Watergate scandal

Temporal relations can be hard to identify

Specific	*before*, *after*
Vague	*in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp	Pim Fortuyn killing, President Nixon resignation
Vague	Watergate scandal

Temporal relations can be hard to identify

Specific	*before*, *after*
Vague	*in the wake of* $\rightarrow$ after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp	Pim Fortuyn killing, President Nixon resignation
Vague	Watergate scandal

Temporal relations can be hard to identify

Specific	*before*, *after*
Vague	*in the wake of* → after? ends? during?

Challenges for temporal QA

Events and states harder to pin down than other named entities

- Variety of syntactic phrase types
- (Almost) no canonical names
- Argument references also syntactically varied

Events and states can be temporally vague

Crisp	Pim Fortuyn killing, President Nixon resignation
Vague	Watergate scandal

Temporal relations can be hard to identify

Specific	*before*, *after*
Vague	*in the wake of* $\rightarrow$ after? ends? during?

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
    - Temporal restriction identification
2. Retrieval: from event/state database
    - Candidate answer events/states
    - If necessary, restricting events/states
3. Temporal event filtering:
    - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture

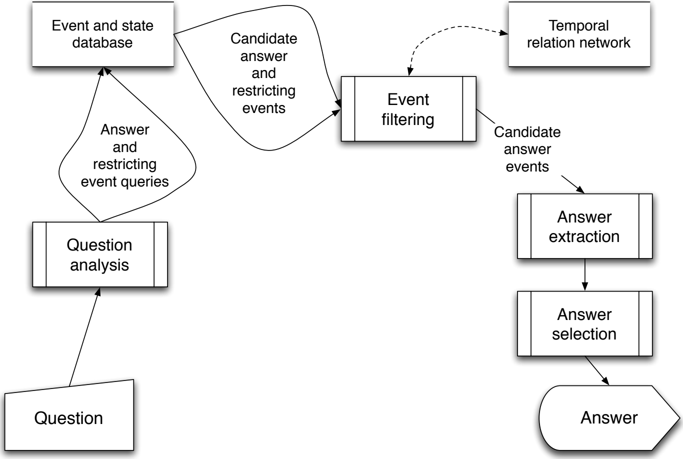
Offline extraction to build knowledge base consisting of:

- Database of events and states, with timestamps
- Temporal relation network for events/states

Question answering process:

1. Question analysis, w/addition of:
   - Temporal restriction identification
2. Retrieval: from event/state database
   - Candidate answer events/states
   - If necessary, restricting events/states
3. Temporal event filtering:
   - Use temporal relation network to filter candidate answer events by temporal restriction
4. Answer extraction and selection, as before

A temporal QA architecture



Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Fuzzy temporal reasoning for filtering

Question: how do we deal with vague events and underspecified relations?

Answer: fuzzy temporal reasoning

- Fuzzy time spans for vague events
- Qualitative temporal relations b/t fuzzy intervals
- Endpoint-based algebra for reasoning
- Network-based representation

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Matching corpus to questions

Question: what kinds of questions do we want to answer?

Answer: it depends

Regardless, the corpora used for extracting answers should match the questions being asked

Matching corpus to questions

Question: what kinds of questions do we want to answer?

Answer: it depends

Regardless, the corpora used for extracting answers should match the questions being asked

Matching corpus to questions

Question: what kinds of questions do we want to answer?

Answer: it depends

Regardless, the corpora used for extracting answers should match the questions being asked

Matching corpus to questions

Question: what kinds of questions do we want to answer?

Answer: it depends

Regardless, the corpora used for extracting answers should match the questions being asked

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE: Squeeze every last drop of information out of each document

QA: Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

Matching corpus to questions

Mismatch:

- Questions: trivia, history
- Corpora: Recent (or not-so-recent) news

Solutions:

- Apply extractors to more relevant corpora: need annotated data for re-training
- Switch methods

Our goal in QA different than the goal in IE:

IE Squeeze every last drop of information out of each document

QA Broadly gather information to answer asked questions

The Web

Full of random statements about events:

- In 1972, the US was reeling from the nightmare of the Vietnam War and the Watergate scandal and the resulting resignation of President Richard Nixon ...
- Watergate was A battle of the Vietnam War. the connection between the Watergate campaign And the Anti-War movement was so closely guarded ..

Perhaps a more reasonable source of information for answering questions about historical trivia?

The Web

Full of random statements about events:

- In 1972, the US was reeling from the nightmare of the Vietnam War and the Watergate scandal and the resulting resignation of President Richard Nixon ...
- Watergate was A battle of the Vietnam War. the connection between the Watergate campaign And the Anti-War movement was so closely guarded ..

Perhaps a more reasonable source of information for answering questions about historical trivia?

The Web

Full of random statements about events:

- In 1972, the US was reeling from the nightmare of the Vietnam War and the Watergate scandal and the resulting resignation of President Richard Nixon ...
- Watergate was A battle of the Vietnam War. the connection between the Watergate campaign And the Anti-War movement was so closely guarded ..

Perhaps a more reasonable source of information for answering questions about historical trivia?

The Web

Full of random statements about events:

- In 1972, the US was reeling from the nightmare of the Vietnam War and the Watergate scandal and the resulting resignation of President Richard Nixon ...
- Watergate was A battle of the Vietnam War. the connection between the Watergate campaign And the Anti-War movement was so closely guarded ..

Perhaps a more reasonable source of information for answering questions about historical trivia?

The Web

Full of random statements about events:

- In 1972, the US was reeling from the nightmare of the Vietnam War and the Watergate scandal and the resulting resignation of President Richard Nixon ...
- Watergate was A battle of the Vietnam War. the connection between the Watergate campaign And the Anti-War movement was so closely guarded ..

Perhaps a more reasonable source of information for answering questions about historical trivia?

What can we extract from the Web?

Given existing events:

- (Fuzzy) time spans for events
- Qualitative relations between events
- New events related to existing events

What can we extract from the Web?

Given existing events:

- (Fuzzy) time spans for events
- Qualitative relations between events
- New events related to existing events

What can we extract from the Web?

Given existing events:

- (Fuzzy) time spans for events
- Qualitative relations between events
- New events related to existing events

What can we extract from the Web?

Given existing events:

- (Fuzzy) time spans for events
- Qualitative relations between events
- New events related to existing events

What can we extract from the Web?

Given existing events:

- (Fuzzy) time spans for events
- Qualitative relations between events
- New events related to existing events

Pattern-based mining

Issues queries to Web search engine using hand-crafted patterns

To find time spans, given `EVENT`:

- `EVENT began on DATE`
- `EVENT lasted until DATE`

To find relations between events (or new, related events), given `EVENT`:

- `EVENT gave way to ALT_EVENT`
- `EVENT took place during ALT_EVENT`

Pattern-based mining

Issues queries to Web search engine using hand-crafted patterns

To find time spans, given `EVENT`:

- `EVENT began on DATE`
- `EVENT lasted until DATE`

To find relations between events (or new, related events), given `EVENT`:

- `EVENT gave way to ALT_EVENT`
- `EVENT took place during ALT_EVENT`

Pattern-based mining

Issues queries to Web search engine using hand-crafted patterns

To find time spans, given `EVENT`:

- `EVENT began on DATE`
- `EVENT lasted until DATE`

To find relations between events (or new, related events), given `EVENT`:

- `EVENT gave way to ALT_EVENT`
- `EVENT took place during ALT_EVENT`

Pattern-based mining

Issues queries to Web search engine using hand-crafted patterns

To find time spans, given `EVENT`:

- `EVENT began on DATE`
- `EVENT lasted until DATE`

To find relations between events (or new, related events), given `EVENT`:

- `EVENT gave way to ALT_EVENT`
- `EVENT took place during ALT_EVENT`

Pattern-based mining

Issues queries to Web search engine using hand-crafted patterns

To find time spans, given `EVENT`:

- `EVENT began on DATE`
- `EVENT lasted until DATE`

To find relations between events (or new, related events), given `EVENT`:

- `EVENT gave way to ALT_EVENT`
- `EVENT took place during ALT_EVENT`

Pattern-based mining

Issues queries to Web search engine using hand-crafted patterns

To find time spans, given `EVENT`:

- `EVENT began on DATE`
- `EVENT lasted until DATE`

To find relations between events (or new, related events), given `EVENT`:

- `EVENT gave way to ALT_EVENT`
- `EVENT took place during ALT_EVENT`

Examples of mined relations

```
<relation type="before" subtype="causal"
 name1="Watergate scandal"
 name2="the resignation of President Richard Nixon">
 <context>In 1972, the US was reeling from the nightmare
  of the Vietnam War and the Watergate scandal and the
  resulting resignation of President Richard Nixon ...
 </context>
</relation>
```

```
<relation type="includes" subtype="part"
 name1="the Vietnam war" name2="Watergate"\>
 <context>Watergate was A battle of the Vietnam War.
  the connection between the Watergate campaign And
  the Anti-War movement was so closely guarded ...
 </context>
</relation>
```

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
  - Coreference with known events
  - Unique reference

Noise

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
  - Coreference with known events
  - Unique reference

Noise

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
  - Coreference with known events
  - Unique reference

Noise

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
    - Coreference with known events
    - Unique reference

Noise

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
  - Coreference with known events
  - Unique reference

Noise

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
  - Coreference with known events
  - Unique reference

Noise

Problems with the Web

Event identification

- Distinguishing references to events from references to other entity types
- Determining the actual reference of an event-referring phrase:
  - Coreference with known events
  - Unique reference

Noise

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Wikipedia as corpus

More relevant than Web for history/trivia QA

- Concentrated collection of factual information

Arguably more reliable than Web

Lots of structure

- Links (w/anchor text)
- Categories
- Lists

Bootstrapping pattern-based mining

Basic idea:

- Grow a seed set of relation instances by extracting patterns from contexts of occurrences of the seed instances in text

Could be used for:

- Relations between times and events
- Relations between events

Bootstrapping pattern-based mining

Basic idea:

- Grow a seed set of relation instances by extracting patterns from contexts of occurrences of the seed instances in text

Could be used for:

- Relations between times and events
- Relations between events

Bootstrapping pattern-based mining

Basic idea:

- Grow a seed set of relation instances by extracting patterns from contexts of occurrences of the seed instances in text

Could be used for:

- Relations between times and events
- Relations between events

Bootstrapping pattern-based mining

Basic idea:

- Grow a seed set of relation instances by extracting patterns from contexts of occurrences of the seed instances in text

Could be used for:

- Relations between times and events
- Relations between events

Bootstrapping pattern-based mining

Basic idea:

- Grow a seed set of relation instances by extracting patterns from contexts of occurrences of the seed instances in text

Could be used for:

- Relations between times and events
- Relations between events

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Pattern bootstrapping algorithm

1. Manually compose a set of seed instances
2. Find all occurrences of each instance in the corpus
3. Extract pattern from each occurrence
4. Rank patterns by reliability
5. Take the best patterns and extract instances produced by the patterns
6. Rank instances by reliability.
7. Use new instances to find more patterns (i.e., repeat from step 2)

Parameters for bootstrapping

Patterns:

- `pre EVENT mid TIMEX post`
- `pre TIMEX mid EVENT post`

Choices for `EVENT` names:

TOKENS Arbitrary sequence of tokens

LINKS Wikipedia links

Choices for `TIMEX`:

- Year
- Full date

Parameters for bootstrapping

Patterns:

- pre EVENT mid TIMEX post
- pre TIMEX mid EVENT post

Choices for EVENT names:

TOKENS Arbitrary sequence of tokens
LINKS Wikipedia links

Choices for TIMEX:

- Year
- Full date

Parameters for bootstrapping

Patterns:

- `pre EVENT mid TIMEX post`
- `pre TIMEX mid EVENT post`

Choices for `EVENT` names:

TOKENS Arbitrary sequence of tokens

LINKS Wikipedia links

Choices for `TIMEX`:

- Year
- Full date

Parameters for bootstrapping

Patterns:

- pre EVENT mid TIMEX post
- pre TIMEX mid EVENT post

Choices for EVENT names:

TOKENS Arbitrary sequence of tokens
LINKS Wikipedia links

Choices for TIMEX:

- Year
- Full date

Parameters for bootstrapping

Reliability measures:

Pattern $r_\pi(p) = \frac{\sum_{i \in I} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\iota(i) \right)}{|I|}$

Instance $r_\iota(i) = \frac{\sum_{p \in P} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\pi(p) \right)}{|P|}$

Number of instances/patterns added

Patterns Round x: top $90 + 10x$ patterns from round $x - 1$

Instances Round x: all instances produced by top $x + 4$ patterns in round x

Parameters for bootstrapping

Reliability measures:

Pattern $r_\pi(p) = \dfrac{\sum_{i \in I} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\iota(i) \right)}{|I|}$

Instance $r_\iota(i) = \dfrac{\sum_{p \in P} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\pi(p) \right)}{|P|}$

Number of instances/patterns added

Patterns Round x: top $90 + 10x$ patterns from round $x - 1$

Instances Round x: all instances produced by top $x + 4$ patterns in round x

Parameters for bootstrapping

Reliability measures:

Pattern $r_\pi(p) = \frac{\sum_{i \in I} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\iota(i) \right)}{|I|}$

Instance $r_\iota(i) = \frac{\sum_{p \in P} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\pi(p) \right)}{|P|}$

Number of instances/patterns added

Patterns Round x: top $90 + 10x$ patterns from round $x - 1$

Instances Round x: all instances produced by top $x + 4$ patterns in round x

Parameters for bootstrapping

Reliability measures:

Pattern $r_\pi(p) = \frac{\sum_{i \in I} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\iota(i) \right)}{|I|}$

Instance $r_\iota(i) = \frac{\sum_{p \in P} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\pi(p) \right)}{|P|}$

Number of instances/patterns added

Patterns Round x: top $90 + 10x$ patterns from round $x - 1$

Instances Round x: all instances produced by top $x + 4$ patterns in round x

Parameters for bootstrapping

Reliability measures:

Pattern $r_\pi(p) = \frac{\sum_{i \in I} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\iota(i) \right)}{|I|}$

Instance $r_\iota(i) = \frac{\sum_{p \in P} \left(\frac{pmi(i,p)}{max_{pmi}} \times r_\pi(p) \right)}{|P|}$

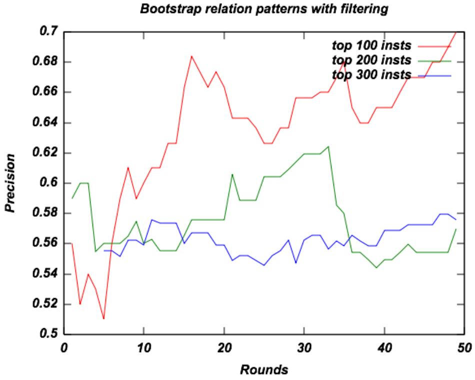
Number of instances/patterns added

Patterns Round x: top $90 + 10x$ patterns from round $x - 1$

Instances Round x: all instances produced by top $x + 4$ patterns in round x

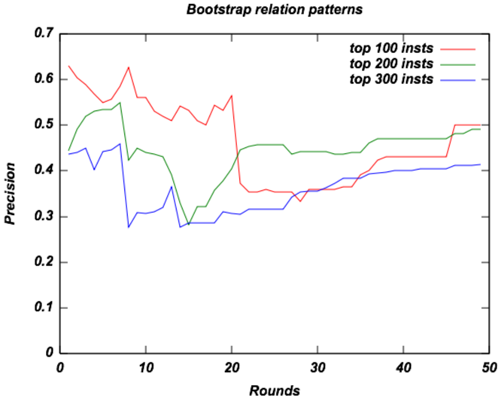
Experiments: Start-of-period

21 seeds (TOKENS event names, year)



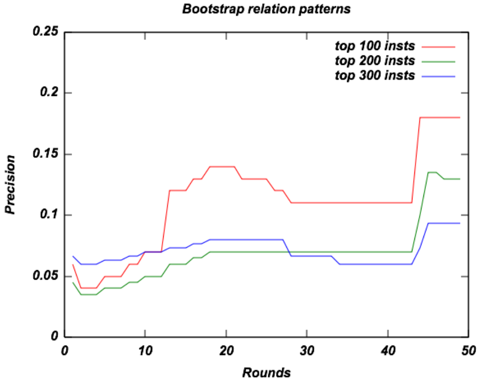
Experiments: Start-of-period

18 seeds (LINKS event names, year)



Experiments: Happened-on

20 seeds (TOKENS event name, full date)



Bootstrapping discussion

Ongoing work—results inconclusive

Restricting event names to Wikipedia links:

- Initial boost followed by decline
- Perhaps adding category information would help

Event-event relations still unexplored

Bootstrapping discussion

Ongoing work—results inconclusive

Restricting event names to Wikipedia links:

- Initial boost followed by decline
- Perhaps adding category information would help

Event-event relations still unexplored

Bootstrapping discussion

Ongoing work—results inconclusive

Restricting event names to Wikipedia links:

- Initial boost followed by decline
- Perhaps adding category information would help

Event-event relations still unexplored

Bootstrapping discussion

Ongoing work—results inconclusive

Restricting event names to Wikipedia links:

- Initial boost followed by decline
- Perhaps adding category information would help

Event-event relations still unexplored

Bootstrapping discussion

Ongoing work—results inconclusive

Restricting event names to Wikipedia links:

- Initial boost followed by decline
- Perhaps adding category information would help

Event-event relations still unexplored

Bootstrapping discussion

Ongoing work—results inconclusive

Restricting event names to Wikipedia links:

- Initial boost followed by decline
- Perhaps adding category information would help

Event-event relations still unexplored

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Recall that one of the reasons for Wikipedia mining is structure

Tried to exploit link/anchor structure for pattern-based mining

What about other structure?

- Categories
- Lists

Wikipedia lists

Wikipedia lists include:

- List of events, births, and deaths for each year
- List of leaders of countries for each year

Year lists: full of events, neatly sorted, with lots of hyperlinks and timestamps

```
* [[August 8]] - [[Watergate scandal]]: US
President [[Richard Nixon]] announces his
resignation (effective [[August 9]])
```

Wikipedia lists

Wikipedia lists include:

- List of events, births, and deaths for each year
- List of leaders of countries for each year

Year lists: full of events, neatly sorted, with lots of hyperlinks and timestamps

```
* [[August 8]] - [[Watergate scandal]]: US President [[Richard Nixon]] announces his resignation (effective [[August 9]])
```

Wikipedia lists

Wikipedia lists include:

- List of events, births, and deaths for each year
- List of leaders of countries for each year

Year lists: full of events, neatly sorted, with lots of hyperlinks and timestamps

```
* [[August 8]] – [[Watergate scandal]]: US
President [[Richard Nixon]] announces his
resignation (effective [[August 9]])
```

Wikipedia lists

Wikipedia lists include:

- List of events, births, and deaths for each year
- List of leaders of countries for each year

Year lists: full of events, neatly sorted, with lots of hyperlinks and timestamps

```
 * [[August 8]] – [[Watergate scandal]]:  US
President [[Richard Nixon]] announces his
resignation (effective [[August 9]])
```

Scraping Wikipedia lists

Very simple structure-based patterns for scraping lists

So far, we have:

- **Events** 58149 events scraped from year entries starting at 1000 AD
- **Leaders** 10400 leaders scraped from lists of state leaders, prime ministers, presidents, governors, and mayors

And we are still exploring lists to scrape . . .

Scraping Wikipedia lists

Very simple structure-based patterns for scraping lists

So far, we have:

Events	58149 events scraped from year entries starting at 1000 AD
Leaders	10400 leaders scraped from lists of state leaders, prime ministers, presidents, governors, and mayors

And we are still exploring lists to scrape . . .

Scraping Wikipedia lists

Very simple structure-based patterns for scraping lists

So far, we have:

Events 58149 events scraped from year entries starting at 1000 AD

Leaders 10400 leaders scraped from lists of state leaders, prime ministers, presidents, governors, and mayors

And we are still exploring lists to scrape . . .

Scraping Wikipedia lists

Very simple structure-based patterns for scraping lists

So far, we have:

Events 58149 events scraped from year entries starting at 1000 AD

Leaders 10400 leaders scraped from lists of state leaders, prime ministers, presidents, governors, and mayors

And we are still exploring lists to scrape ...

Scraping Wikipedia lists

Very simple structure-based patterns for scraping lists

So far, we have:

Events 58149 events scraped from year entries starting at 1000 AD

Leaders 10400 leaders scraped from lists of state leaders, prime ministers, presidents, governors, and mayors

And we are still exploring lists to scrape . . .

Scraping Wikipedia lists

Very simple structure-based patterns for scraping lists

So far, we have:

Events 58149 events scraped from year entries starting at 1000 AD

Leaders 10400 leaders scraped from lists of state leaders, prime ministers, presidents, governors, and mayors

And we are still exploring lists to scrape . . .

Building a knowledge base for historical QA

Scraping approach seems most promising for building foundation of knowledge base for historical QA

Other approaches could be used to fill out knowledge base

Storage and retrieval: events and leaders stored and indexed as individual Lucene documents, for standard text retrieval

Building a knowledge base for historical QA

Scraping approach seems most promising for building foundation of knowledge base for historical QA

Other approaches could be used to fill out knowledge base

Storage and retrieval: events and leaders stored and indexed as individual Lucene documents, for standard text retrieval

Building a knowledge base for historical QA

Scraping approach seems most promising for building foundation of knowledge base for historical QA

Other approaches could be used to fill out knowledge base

Storage and retrieval: events and leaders stored and indexed as individual Lucene documents, for standard text retrieval

Building a knowledge base for historical QA

Scraping approach seems most promising for building foundation of knowledge base for historical QA

Other approaches could be used to fill out knowledge base

Storage and retrieval: events and leaders stored and indexed as individual Lucene documents, for standard text retrieval

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Question analysis:

Answer type Person

Answer event `Dutch Prime Minister`

Restricting event `Pim Fortuyn shot`

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Question analysis:

Answer type Person

Answer event `Dutch Prime Minister`

Restricting event `Pim Fortuyn shot`

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Question analysis:

Answer type Person

Answer event `Dutch Prime Minister`

Restricting event `Pim Fortuyn shot`

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Question analysis:

Answer type Person

Answer event `Dutch Prime Minister`

Restricting event `Pim Fortuyn shot`

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Question analysis:

Answer type Person

Answer event `Dutch Prime Minister`

Restricting event `Pim Fortuyn shot`

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Retrieval:

- Queries `Dutch Prime Minster` and `Pim Fortuyn shot` yield no results
- Problem: event descriptions too short for lexical matching
- Possible solutions:
  - Lexical resources: map `Dutch` to `Netherlands` and `shot` to `killed`, `died`
  - Mixture models: interpolate text from linked entries into event descriptions

For now, assume necessary lexical resources

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Candidate answer events:

1901–1905 prime_minister(Abraham Kuyper, Netherlands)
1872–1874 prime_minister(Gerrit de Vries, Netherlands)
...
1994-08-22–2002-07-22 prime_minister(Wim Kok, Netherlands)
2004–present prime_minister(Etienne Ys, Netherlands Antilles)

Candidate restricting events:

2002-05-06 Died - Pim Fortuyn, Dutch politician (assassinated)
2002-05-06 In the Netherlands, politician Pim Fortuyn is killed by Volkert van der Graaf.

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Candidate answer events:

1901–1905	prime_minister(Abraham Kuyper, Netherlands)
1872–1874	prime_minister(Gerrit de Vries, Netherlands)
	...
1994-08-22–2002-07-22	prime_minister(Wim Kok, Netherlands)
2004–present	prime_minister(Etienne Ys, Netherlands Antilles)

Candidate restricting events:

2002-05-06	Died - Pim Fortuyn, Dutch politician (assassinated)
2002-05-06	In the Netherlands, politician Pim Fortuyn is killed by Volkert van der Graaf.

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Candidate answer events:

1901–1905	prime_minister(Abraham Kuyper, Netherlands)
1872–1874	prime_minister(Gerrit de Vries, Netherlands)
	...
1994-08-22–2002-07-22	prime_minister(Wim Kok, Netherlands)
2004–present	prime_minister(Etienne Ys, Netherlands Antilles)

Candidate restricting events:

2002-05-06	Died - Pim Fortuyn, Dutch politician (assassinated)
2002-05-06	In the Netherlands, politician Pim Fortuyn is killed by Volkert van der Graaf.

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Candidate answer events:

1901–1905	prime_minister(Abraham Kuyper, Netherlands)
1872–1874	prime_minister(Gerrit de Vries, Netherlands)
	...
1994-08-22–2002-07-22	prime_minister(Wim Kok, Netherlands)
2004–present	prime_minister(Etienne Ys, Netherlands Antilles)

Candidate restricting events:

2002-05-06	Died - Pim Fortuyn, Dutch politician (assassinated)
2002-05-06	In the Netherlands, politician Pim Fortuyn is killed by Volkert van der Graaf.

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Candidate answer events:

1901–1905 prime_minister(Abraham Kuyper, Netherlands)
1872–1874 prime_minister(Gerrit de Vries, Netherlands)
...
1994-08-22–2002-07-22 prime_minister(Wim Kok, Netherlands)
2004–present prime_minister(Etienne Ys, Netherlands Antilles)

Candidate restricting events:

2002-05-06 Died - Pim Fortuyn, Dutch politician (assassinated)
2002-05-06 In the Netherlands, politician Pim Fortuyn is killed by Volkert van der Graaf.

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Temporal event filtering:

1994-08-22–2002-07-22 prime_minister(Wim Kok, Netherlands)

Answer extraction:

- Wim Kok

Answer selection:

- Wim Kok

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Temporal event filtering:

1994-08-22–2002-07-22 prime_minister(Wim Kok, Netherlands)

Answer extraction:

- Wim Kok

Answer selection:

- Wim Kok

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Temporal event filtering:

1994-08-22–2002-07-22 prime_minister(Wim Kok, Netherlands)

Answer extraction:

- Wim Kok

Answer selection:

- Wim Kok

Example: Who was the Dutch Prime Minister when Pim Fortuyn was shot?

Temporal event filtering:

1994-08-22–2002-07-22 prime_minister(Wim Kok, Netherlands)

Answer extraction:

- Wim Kok

Answer selection:

- Wim Kok

Outline

1. Temporal QA
2. Web mining
3. Wikipedia mining
   - Bootstrapping pattern-based mining
   - Scraping Wikipedia lists
4. Looking forward

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions

Still to be done

Knowledge base building

- More scraping
- Experiment with mining event-event relations
- Manual annotation for possible IE
- Experiment with mixture models for indexing

Temporal question-specific analysis module

Most important: Find sources of temporal questions

- QA research is best guided by actual questions