Temporal bounds come from below document level

The goal is to retrieve as much temporal bounding information as possible

Document level timestamps are no good

Searching for "Tabas earthquake" (a 1978 event in Iran) gives document dates having these kinds of temporal clusters:

Why are the temporal spikes "wrong"?

News articles contain the phrase, but also tend to refer to similar prior events for context. In fact, a typical news article has a complex temporal structure:

What information can help determine temporal bounds?

Events – exploded, elected
Temporal expressions – May, 2004, next Tuesday
Connections between them, within and across documents

Any information that contributed to determining a temporal bound is a:

temporal description

To perform information retrieval for temporal bounding, retrieve as many temporal descriptions as possible for a given assertion

How hard could it be?

Preliminary temporal description retrieval experiment
This problem is less interesting if it's easy to solve. Is it?

Assertions: available as gold-standard, from Wikipedia infoboxes

Dataset: 1.8M documents of various genre, publication dates spread over a ten year span

Evaluation: Compare with human-marked temporal descriptions

IR system: Lucene

We compare two systems

Version one: no additional information, use assertion as query

Version two: annotate document for temporal expressions, prefer documents where query occurs near temporal expression

Evaluation

Perform evaluation with metric also used in IR for QA:

Coverage – the proportion of assertions that have at least one relevant TD-bearing document in the results set.

No performance gain – actually a slight drop. The problem is non-trivial

Problem summary

We often retrieve and return results with multiple assertions, in cases where only one can be true at a time

To remedy this, we need to determine temporal bounds of information.

This can be achieved through finding descriptions of temporal bounds in a document collection – a non-trivial task.

References

