Open Knowledge Representation for Textual Information

Ido Dagan, Bar-Ilan University
DMBI 2017
Motivation
A Perspective: Structured Knowledge Graphs

Freebase knowledge graph
Structured Knowledge Graphs (KG)

- Encode selected facts about the world
- As propositions - asserted relations (predicates) between entities

- Allow structured presentation, querying and inference
- Relations and entities are pre-specified
  - Practically inventing a new “language” when defining a graph/ontology schema
Our Motivation: Open NL-based KR (OKR)

Research questions:

- How can we capture **unbounded** information in collections of text?
- How can we communicate it efficiently to users, or applications?

Proposed answers:

- An “open” **NL-based** representation
  - Utilize the available human language, vs. inventing a new formal one
- Enables effective text exploration at **statement/fact** level
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Turkey forces a **plane from Russia** to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

**Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia** to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says **plane from Moscow** forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger **plane enroute from Moscow** land in Ankara.
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

 Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Characteristic 1: Redundancy

Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Background:
Derive OKR (mostly) from known component tasks
Single Sentence Representations (1)

- **SRL** - annotating predicate-argument structure of sentences.

  *Danny arrived to the restaurant to meet Sarah*

- **QA-SRL** - an argument role is a natural language question which the argument answers

- **Open IE** - isolates “atomic” propositions as predicate-argument tuples
  
  *(Danny, arrived, at restaurant, to meet Sarah)*
  
  *(Danny, meet, Sarah)*
Cross-document Pred-arg Alignment/Coreference

- Roth and Frank (2012), PARMA (2013), ECB+ (2014) - Cross-document alignment of events and entities

  T1: On Monday Lindsay Lohan checked into rehab in Malibu, California after car crash.
  T2: Ms. Lohan entered a treatment facility.
  T3: The 26 year old actress checked into Betty Ford Center.

- Each of these efforts is limited/partial in certain respects
- No consolidation
Coreference-based inference

- Lexical substitution of co-referring elements
  - *Turkey intercepted a Syrian jet. The plane was carrying ammunition*
    - *A Syrian jet was carrying ammunition*

- Used in question answering, Recognizing Textual Entailment, etc.

- In formal semantics: DRT (Kamp 1988, Kamp et al., 2011)
  - A proposition applies to all co-referring (equal) entities
Our Positioning and Novelty

- Rely on extended Open-IE extractions and entity/predicate coreference

- **Merge pred-arg extractions** into a consolidated structure
  - Based on coreference relations

- **Trace redundancies** - through entailment relations
Graph Specification & Construction
1. They bought... the Pasqua chain of 56 coffee stores...

2. Purchase of Pasqua Coffee that year added 56 outlets in San Francisco.


4. Starbucks acquired San Francisco-based Pasqua Coffee, adding 56 outlets to its network.
Starbucks acquired SF-based Pasqua Coffee, adding 56 outlets to its network.
Starbucks acquired SF-based Pasqua Coffee, adding 56 outlets to its network.
Consolidation - Element Coreference

1. They bought the Pasqua chain who has 56 stores what coffee.

2. Purchase Pasqua Coffee that year 56 outlets added San Francisco.

3. Acquired Starbucks Pasqua Coffee when has 56 outlets 1998.

4. Adding Starbucks 56 outlets network acquired based Pasqua Coffee SF.
Consolidation - Element Coreference

1. They bought the Pasqua chain who has 56 outlets adding what to what.

2. Purchase added Pasqua Coffee that year 56 outlets San Francisco.

3. Starbucks acquired Pasqua Coffee 1998 when has 56 outlets what U.S.

4. Adding Starbucks acquired Pasqua Coffee has network 56 outlets based SF.
The Pasqua chain bought coffee from the Pasqua chain in 1998. Starbucks acquired Pasqua Coffee in 1998, when the company had 56 outlets in the U.S. Starbucks added 56 outlets to its network in San Francisco that year. The company has 56 outlets based in SF.
Boldness signifies frequency.
Dataset and Baselines
News-Related Tweets

- OKR Annotation of 1257 news-related tweets from 27 events clusters.
  - Collected from the Twitter Event Detection Dataset (McMinn et al., 2013)

- Automated baseline system under development
Applications
Enhance Cross-text Applications

● Text summarization, question answering, text analytics/mining, ...
● Currently, some such applications perform partial & application-specific cross-text consolidation

Our Proposal:
  ○ Add cross-text information consolidation to the NLP pipeline,
  ○ instead of having to process each sentence separately
Interactive Text Exploration

● "What is said here?"
  ○ Dynamic: by salience, detail level, length
    ■ F-16 vs. Turkish army vs. Turkey
  ○ Hierarchical
  ○ Incremental
  ○ Bullet style vs. continuous text

● Improve over (static) abstractive multi-document summarization
Prototype Implementation

- A prototype exploration system was implemented for news tweets
  - Dynamic bullet summary of events (~100 tweets)

- Tweets are consolidated to a few abstractive summary sentences
  - Cover all the information without redundancy
  - Emphasize fact importance
  - Allow drill down to details
Wisconsin_shooting
Event Summary

22.10.12 01:07 3 Tweets a number of people dead all were women

21.10.12 10:55 17 Tweets suspected gunman confirmed dead

21.10.12 10:11 10 Tweets suspected gunman still at large after at least seven injured

21.10.12 09:52 59 Tweets salon outside milwaukee
Support Tweets

a number of people dead all were women

suspected gunman confirmed dead

suspected gunman still at large after at least seven injured

salon outside milwaukee

Total knowledge covered by opened sentences

Sentence knowledge contribution

Less informative sentences

Open Knowledge Representation

Wisconsin_shooting
Event Summary

Timeline

22.10.12 01:07 3 Tweets

21.10.12 10:55 17 Tweets

21.10.12 10:11 10 Tweets

21.10.12 09:52 59 Tweets

56% Knowledge Presented

9.1% Knowledge
Concept expansion: complementary (co-referring) information about a concept:

Pressing a concept highlights the mentions throughout the summary:
Preliminary Usability Study

<table>
<thead>
<tr>
<th>SUS Question</th>
<th>Avg Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think that I would like to use this system frequently.</td>
<td>3.83</td>
</tr>
<tr>
<td>I thought the system was easy to use.</td>
<td>3.33</td>
</tr>
<tr>
<td>I found the various functions in this system were well integrated.</td>
<td>3.83</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use this system very quickly.</td>
<td>3.5</td>
</tr>
<tr>
<td>I felt very confident using the system.</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Table 1: System Usability Questions - average answer score over the six users on a scale of 1 to 5.

<table>
<thead>
<tr>
<th>User</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS Score</td>
<td>70</td>
<td>80</td>
<td>95</td>
<td>72.5</td>
<td>82.5</td>
<td>27.5</td>
</tr>
</tbody>
</table>

Table 2: System Usability Scores

<table>
<thead>
<tr>
<th>Dimension</th>
<th>$B_t$</th>
<th>$B_g$</th>
<th>Main</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>2.1</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Knowledge Exploration</td>
<td>2.0</td>
<td>1.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>2.0</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>2.5</td>
<td>2.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Ease of Learning</td>
<td>2.7</td>
<td>2.5</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Table 3: USE questionnaire dimensions score comparison of the three system interfaces on a scale of 1 to 3.
Summary

- **Novel Open Knowledge Representation**
  - Rely on extended Open-IE *extractions* and *entity/predicate coreference*
  - Consolidated structure
  - Trace redundancies
  - Gold-Standard Dataset

- **Interactive text exploration**
  - Initial assessment in standard usability study