Mark Spitz

From Wikipedia, the free encyclopedia

Mark Andrew Spitz (born February 10, 1950) is an American former competition swimmer, nine-time Olympic champion, and former world record-holder in seven events. He won seven gold medals at the 1972 Summer Olympics in Munich, an achievement surpassed only by Michael Phelps, who won eight golds at the 2008 Summer Olympics in Beijing. Spitz set new world records in all seven events in which he competed in 1972, an achievement that still stands. Since the year 1900, no other swimmer has gained so great a percentage of all the medals awarded for Olympic events held in a single Games.
Mark Spitz
From Wikipedia, the free encyclopedia

Mark Andrew Spitz (born February 10, 1950) is an American former competition swimmer, nine-time Olympic champion, and former world record-holder in seven events. He won seven gold medals at the 1972 Summer Olympics in Munich, an achievement surpassed only by Michael Phelps, who won eight golds at the 2008 Summer Olympics in Beijing. Spitz set new world records in all seven events in which he competed in 1972, an achievement that still stands. Since the year 1900, no other swimmer has gained so great a percentage of all the medals awarded for Olympic events held in a single Games.

Olympia Schwimmhalle
From Wikipedia, the free encyclopedia

The Olympia Schwimmhalle is an aquatics centre located in the Olympiapark in Munich, Germany. It hosted the swimming, diving, water polo, and the swimming part of the modern pentathlon events at the 1972 Summer Olympics. At the 1972 Olympics, the stadium had a 9000-seat capacity which was reduced to 1,500 soon after. During the 1972 Olympics, the Olympic Records in all 29 Olympic swimming events were broken as well as the World Records in 20 events.[citation needed]

The Schwimmhalle is unique for its roof construction which is a lightweight stressed-skin structure. This curved structure bears loads through tension only, not compression. The double curvature in the roof design is what provides support which is further stabilized through pretensioned guy wires.

The Olympia Schwimmhalle is where swimmer Mark Spitz broke the record for most individual gold medals won in a single Olympics with seven gold medals. This record was not surpassed until fellow swimmer Michael Phelps won eight gold medals at the 2008 Summer Olympics in Beijing.

1972 Summer Olympics
From Wikipedia, the free encyclopedia

The 1972 Summer Olympics (German: Olympische Sommerspiele 1972), officially known as the Games of the XX Olympiad, was an international multi-sport event held in Munich, West Germany, from August 26 to September 11, 1972.

Steve Genter
From Wikipedia, the free encyclopedia

Robert Steven Genter (born January 4, 1951) is an American former competition swimmer and three-time Olympic medalist. He was freestyle specialist who earned a gold medal as a member of the winning U.S. team in the 4×200-meter freestyle relay at the 1972 Summer Olympics in Munich, Germany. He also won silver medals in the 200-meter and 400-meter freestyle events.

Swimming at the 1972 Summer Olympics
From Wikipedia, the free encyclopedia

The 1972 Summer Olympics were held in Munich, West Germany, 29 events in swimming were contested. There was a total of 532 participants from 52 countries competing.
July 4, 2016
Runner Boris Berian Goes From
McDonald's Employee to a Symbol of
Athletes' Rights

An 800-meter runner, Berian found himself embroiled in a legal battle with the athletic-apparel behemoth Nike when he attempted to change sponsors.

By RISHI LUNGAWI

July 4, 2016
Vashti Cunningham, Daughter of an N.F.L.
Improvisor, Is Raising the Bar

The 18-year-old Cunningham, the offspring of the former N.F.L. quarterback Randall Cunningham, will compete at the Rio Olympics in the high jump.

By RISHI LUNGAWI

July 3, 2016
SPORTS BRIEFING
Russia Appeals Doping Ban to International Sports Court

The appeal will be heard by the Court of Arbitration for Sport on July 19. The Russian Olympic Committee said it would extend the deadline for application to the Olympics if it succeeds.
**Definition: Event**

"Something that happens at a given place and time between a group of actors." [CSG⁺02]
Motivation

Definition: Event

“Something that happens at a given place and time between a group of actors.”

[CSG+02]

For large document collections, how can we...

- obtain events from unstructured text?
- identify connections across documents?
- support ad-hoc event search?
Graph Extraction from Unstructured Text
Graph Extraction from Unstructured Text

Terms over LOAD: Named Entities for Cross-Document Event Extraction

Andreas Spitz
Graph Extraction from Unstructured Text
Graph Extraction from Unstructured Text

- Terms over LOAD: Named Entities for Cross-Document Event Extraction
- Andreas Spitz
Graph Extraction from Unstructured Text

Terms over LOAD: Named Entities for Cross-Document Event Extraction

Andreas Spitz
Edge Weight Generation

For edges \((x, y)\) for which \(y\) is a page or sentence, count only (co-) occurrences:

\[
\omega(x, y) = \begin{cases} 
1 & \text{if } y \text{ contains } x \\
0 & \text{otherwise}
\end{cases}
\]
Edge Weight Generation

For edges \((x, y)\) for which \(y\) is a page or sentence, count only (co-) occurrences:

\[
\omega(x, y) = \begin{cases} 
1 & \text{if } y \text{ contains } x \\
0 & \text{otherwise}
\end{cases}
\]

For edges \((x, y)\) between entity types and terms, aggregate co-occurrence instances \(I\): sum over similarities derived from sentence distances \(s\).

\[
\omega(x, y) := \sum_{i \in I} \exp(-s(x, y, i))
\]
LOADing Wikipedia

For the entire English Wikipedia (∼ 4.5M articles with annotations):

- use only **unstructured** text.
- exclude pages of lists.
- exclude info boxes.
- exclude references.

Extract named entities with:

- Stanford NER for **locations**, **organizations** and **actors** [FGM05]
- Heideltime for **dates** [SG13]
### Wikipedia LOAD Graph

<table>
<thead>
<tr>
<th>edges</th>
<th>LOC</th>
<th>ORG</th>
<th>ACT</th>
<th>DAT</th>
<th>TER</th>
<th>SEN</th>
<th>PAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORG</td>
<td>91</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT</td>
<td>276</td>
<td>106</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>83</td>
<td>46</td>
<td>128</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TER</td>
<td>183</td>
<td>94</td>
<td>317</td>
<td>57</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEN</td>
<td>71</td>
<td>21</td>
<td>84</td>
<td>38</td>
<td>412</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PAG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>54</td>
</tr>
</tbody>
</table>

| nodes | 2.7 | 3.4 | 7.1 | 0.2 | 4.9 | 53.5 | 4.5 |

Number of edges and nodes (in millions) of the LOAD graph of the English Wikipedia. \( \sim 2B \) edges and \( \sim 76M \) nodes in total.
Single Entity Queries

How can we rank nodes in one set $Y$ by their neighbours in set $X$?
Adapt $tf-idf$ scores to the graph [RV13]!

- Term frequency: edge weights
  
  \[ tf(x, y) \approx \omega(x, y) \]

- Inverse document frequency: number of neighbours
  
  \[ idf(x) \approx \frac{|Y|}{\text{deg}_Y(x)} \]

\[ r(x, y) \approx \omega(x, y) \log \frac{|Y|}{\text{deg}_Y(x)} \]
Single Entity Queries

How can we rank nodes in one set $Y$ by their neighbours in set $X$? Adapt \textit{tf-idf} scores to the graph \cite{RV13}!

- Term frequency: edge weights
  $$tf(x, y) \approx \omega(x, y)$$

- Inverse document frequency: number of neighbours
  $$idf(x) \approx \frac{|Y|}{\text{deg}_Y(x)}$$

$$r(x, y) \approx \omega(x, y) \log \frac{|Y|}{\text{deg}_Y(x)}$$

\[ \langle \text{LOC} : (ACT, \text{Mark Spitz}) \rangle \]

<table>
<thead>
<tr>
<th>location</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>munich</td>
<td>1.00000</td>
</tr>
<tr>
<td>us</td>
<td>0.70651</td>
</tr>
<tr>
<td>states</td>
<td>0.49010</td>
</tr>
<tr>
<td>united states</td>
<td>0.46918</td>
</tr>
</tbody>
</table>

Query: \[ \langle Y : (X, \text{value}) \rangle \]
Multi-Entity Queries

How can we rank nodes in $Y$ by neighbours in multiple sets $X^n$?

Combine individual set scores:

$$r(\vec{x}, y) := \frac{1}{n} \eta(\vec{x}, y) \sum_{i=1}^{n} r(x_i, y)$$

Where $M$ is the adjacency matrix of the graph.
Multi-Entity Queries

How can we rank nodes in $Y$ by neighbours in multiple sets $X^n$? Combine individual set scores:

$$r(\vec{x}, y) := \frac{1}{n} \eta(\vec{x}, y) \sum_{i=1}^{n} r(x_i, y)$$

Ensure triangular cohesion when combining results:

$$\eta(\vec{x}, y) := \begin{cases} 1 & \text{if } \sum_{i=1}^{n} \sum_{j>i}^{n} M_{yx_i} M_{yx_j} > 1 \\ 0 & \text{otherwise} \end{cases}$$

Where $M$ is the adjacency matrix of the graph.
## Multi-Entity Query Examples

\[
\langle DAT : (ACT, \text{Mark Spitz}), (LOC, \text{Munich}) \rangle
\]

<table>
<thead>
<tr>
<th>date</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972-08-29</td>
<td>0.50851</td>
</tr>
<tr>
<td>1972-08-31</td>
<td>0.48217</td>
</tr>
<tr>
<td>1972-09-05</td>
<td>0.22738</td>
</tr>
<tr>
<td>1947-03-10</td>
<td>0.10511</td>
</tr>
<tr>
<td>2006-09-07</td>
<td>0.09226</td>
</tr>
</tbody>
</table>
## Multi-Entity Query Examples

### Example 1: \( \langle DAT : (ACT, \text{Mark Spitz}), (LOC, \text{Munich}) \rangle \)

<table>
<thead>
<tr>
<th>date</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972-08-29</td>
<td>0.50851</td>
</tr>
<tr>
<td>1972-08-31</td>
<td>0.48217</td>
</tr>
<tr>
<td>1972-09-05</td>
<td>0.22738</td>
</tr>
<tr>
<td>1947-03-10</td>
<td>0.10511</td>
</tr>
<tr>
<td>2006-09-07</td>
<td>0.09226</td>
</tr>
</tbody>
</table>

### Example 2: \( \langle TER : (ACT, \text{Mark Spitz}), (LOC, \text{Munich}), (DAT, 1972) \rangle \)

<table>
<thead>
<tr>
<th>term</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>olymp</td>
<td>0.89630</td>
</tr>
<tr>
<td>medal</td>
<td>0.54205</td>
</tr>
<tr>
<td>gold</td>
<td>0.43211</td>
</tr>
<tr>
<td>won</td>
<td>0.38904</td>
</tr>
<tr>
<td>record</td>
<td>0.34548</td>
</tr>
</tbody>
</table>
Summarization: Sentence Queries

How can sentences in $S$ be used to describe combinations of entities in $X^n$?

Find a sentence that contains them:

$$r(\vec{x}, s) := \sum_{i=1}^{n} M_{sx_i}$$

Mark Spitz of the United States had a spectacular run, lining up for seven events, winning seven Olympic titles and setting seven world records. 
Summarization: Sentence Queries

How can sentences in $S$ be used to describe combinations of entities in $X^n$?

Find a sentence that contains them:

$$r(\vec{x}, s) := \sum_{i=1}^{n} M_{sx_i}$$

$$\langle SEN : (ACT, \text{Mark Spitz}) \rangle$$

Mark Spitz of the United States had a spectacular run, lining up for seven events, winning seven Olympic titles and setting seven world records.
Entity Linking: Document Queries

Since we created the LOAD graph from Wikipedia, can we link entities in $X^n$ to pages $P$?

Use sentences to find the page that contains them most frequently:

$$r(\vec{x}, p) := \sum_{s \in S} \sum_{i=1}^{n} M_{sx_i} M_{sp}$$
Entity Linking: Document Queries

Since we created the LOAD graph from Wikipedia, can we link entities in $X^n$ to pages $P$?

Use sentences to find the page that contains them most frequently:

$$r(\vec{x}, p) := \sum_{s \in S} \sum_{i=1}^{n} M_{sx_i} M_{sp}$$

$\langle PAG : (ACT, Mark Spitz) \rangle$

Wiki page ID 66265: Mark Spitz
Intuition:

- Events correspond to triangular structures in the network
- Participating entities can be used to complete events
Query Answering Speed

Asymptotic complexity of entity queries: $O(\deg_X(y) \deg_Y(x))$
Historic Event Evaluation Data

Evaluation data set from a “This Day in History” website [Gui95]

- old enough to not contain Wikipedia data
- exactly one date per sentence
- 500 hand-annotated historic events
- example: Ernest Hemingway, Red Cross volunteer, wounded in Italy on 1918-07-08.
Evaluation on Historic Event Data

Retrieving Dates of Historic Events

<table>
<thead>
<tr>
<th>fraction of included dates</th>
<th>method</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOADTsq</td>
<td></td>
</tr>
<tr>
<td>LOADsq</td>
<td></td>
</tr>
<tr>
<td>LOAD</td>
<td></td>
</tr>
<tr>
<td>BASEw</td>
<td></td>
</tr>
</tbody>
</table>

maximum rank

0.0
0.1
0.2
0.3
0 10 20 30 40 50 60 70 80 90 100
NER based on Wikipedia & Wikidata
Summary

Ongoing work:

- online search and query interface for Wikipedia
- streaming model for online news
- inclusion of parts-of-speech

LOAD summary:

- fast entity and event exploration
- can support most entity-related IE tasks
- can be extended to any kind of entity
- scalable and fast
- language-agnostic with entity linking
Summary

Ongoing work:
- online search and query interface for Wikipedia
- streaming model for online news
- inclusion of parts-of-speech

LOAD summary:
- fast entity and event exploration
- can support most entity-related IE tasks
- can be extended to any kind of entity
- scalable and fast
- language-agnostic with entity linking

LOAD your data before you do entity-based analyses.
Available for download:

- Wikipedia LOAD network (Stanford NER)
- Wikipedia LOAD network (Wikidata)
- Code for generating LOAD networks
- Code for LOAD query interface

http://dbs.ifi.uni-heidelberg.de/index.php?id=load
Available for download:
- Wikipedia LOAD network (Stanford NER)
- Wikipedia LOAD network (Wikidata)
- Code for generating LOAD networks
- Code for LOAD query interface

http://dbs.ifi.uni-heidelberg.de/index.php?id=load
Bibliography

Christopher Cieri, Stephanie Strassel, David Graff, Nii Martey, Kara Rennert, and Mark Liberman.
Corpora for topic detection and tracking.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning.
Incorporating non-local information into information extraction systems by Gibbs sampling.
In *ACL*, 2005.

Robert A Guisepi.
History world: On this day in history, 1995.

François Rousseau and Michalis Vazirgiannis.
Graph-of-word and TW-IDF: new approach to ad hoc IR.
In *CIKM*, 2013.

Jannik Strötgen and Michael Gertz.
Multilingual and cross-domain temporal tagging.