TimeMachine

TIMELINE GENERATION FOR KNOWLEDGE-BASE ENTITIES

Tim Althoff (Stanford CS), Xin Luna Dong, Kevin Murphy, Safa Alai, Van Dang, Wei Zhang (Google)
Learning about new topics is hard

Really common scenario!

Even learning about a simple topic like a person!
Web Search Results

Robert Downey Jr. - Wikipedia, the free encyclopedia
Robert Downey Jr. (born April 4, 1965) is an American actor whose career has included critical and popular success in his youth, followed by a period of ...
Robert Downey, Sr. - Robert Downey Jr. filmography - Susan Downey - Chaplin

In the news
How Much Did Robert Downey Jr Make for 'Avengers 2' & 'Captain America: Civil War'? /FILM - 10 hours ago
A couple years ago, Robert Downey Jr. hinted that he might be done with superhero movies ...
Robert Downey Jr. Sent Chris Pratt The Greatest 'Jurassic World' Meme
Huffington Post - 2 days ago
Robert Downey Jr. Makes A Starling Amount Of Money From Marvel
Cinema Blend - 3 hours ago

More news for robert downey jr

Robert Downey Jr. - IMDb
www.imdb.com/name/nm0000375 - Internet Movie Database
Robert Downey Jr., Actor: The Avengers. Robert Downey Jr. has the most respected actors in Hollywood. With an amazing list of cre...
Structured Information

Manual curation

Bias towards recent blockbusters

Robert Downey Jr.

Actor

Robert John Downey, Jr. is an American actor whose career has included critical and popular success in his youth, followed by a period of substance abuse and legal troubles, and a resurgence of commercial success in middle age. Wikipedia

Born: April 4, 1965 (age 50), Manhattan, New York City, NY
Height: 5' 9" (1.74 m)
Spouse: Susan Downey (m. 2005), Deborah Falconer (m. 1982–2004)
Parents: Elsie Ford, Robert Downey, Sr.
Children: Indio Falconer Downey, Exton Elias Downey, Avri Roel Downey

Missing context of relationship
Our System: TimeMachine

“RDJr was award-nominated for Ally McBeal in 2000.”

Better overview through timeline visualization

Allows for further exploration such as zooming
Timeline Generation

Problem Definition

Entity

Timeline

- Timeline needs to be rendered on device
- Generation must be fast (zoom + interaction)
Timeline Generation Approach

1. Event Generation

Entity → List of events

2. Event Selection

List of events → Timeline

- The Avengers
  - Samuel L. Jackson
- Doctor Who
  - Peter Capaldi
- The Matrix
  - Keanu Reeves (1999–2008)
  - Laurence Fishburne
- The Dark Knight
  - Heath Ledger
- The Avengers
  - Mark Ruffalo

- Timeline:
  - (2012) The Avengers
  - (2015) Ant-Man
  - (2016) Captain America: Civil War
  - (2016) Doctor Strange
  - (2016) Batman v Superman: Dawn of Justice
  - (2016) Deadpool
  - (2017) Spider-Man: Homecoming
  - (2017) Logan
  - (2017) Wonder Woman
  - (2017) The Last Jedi
  - (2018) Black Panther
  - (2018) Avengers: Infinity War
  - (2018) Aquaman
  - (2019) Captain Marvel
  - (2019) Joker
  - (2019) Doctor Strange in the Multiverse of Madness
  - (2020) The Mandalorian
  - (2020) The Trial of the Chicago 7
  - (2020) The Irishman
  - (2021) Shang-Chi and the Legend of the Ten Rings
  - (2021) Black Widow
  - (2021) The Eternals
  - (2021) Spider-Man: No Way Home
  - (2021) Kantai Collection
  - (2021) The Matrix Resurrections
  - (2022) Moon Knight
  - (2022) The Suicide Squad
  - (2022) The Flash
  - (2022) Ant-Man and the Wasp: Quantumania
## List of Events

<table>
<thead>
<tr>
<th>Subject</th>
<th>Related Entity</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. D. Jr</td>
<td>Robert D. Sr</td>
<td>1988</td>
<td>In movie directed by</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jon Bon Jovi</td>
<td></td>
<td>2004</td>
<td>TV show app. with</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Susan Downey</td>
<td></td>
<td>2005</td>
<td>Got married</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron Man</td>
<td></td>
<td>2008</td>
<td>Award for movie</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avengers</td>
<td></td>
<td>2010</td>
<td>Acted in movie</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Event Generation
Candidate Event Generation

Robert Downey Jr.  

DoB  

April 4, 1965

Subject

Related Entity

Timestamp
## Resulting Candidate Events

<table>
<thead>
<tr>
<th>Subject</th>
<th>Related Entity</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. D. Jr</td>
<td>-</td>
<td>1965</td>
<td>Was born</td>
</tr>
<tr>
<td>The Avengers</td>
<td></td>
<td>2012</td>
<td>Acted in movie</td>
</tr>
<tr>
<td>Samuel L Jackson</td>
<td></td>
<td>2012</td>
<td>Co-starred in The Avengers</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

737 events total
Event Filtering

• Some non-informative events: any American → nationality: USA → founded 1776

• **Frequency Filter:** events commonly associated with a large number of subjects are unlikely to be interesting (like IDF) (no “nationality → founded”)

• **Existence Filter:** filter out events before entity begins to exist (no “parent → DOB”)
Event Filtering Evaluation

- 87% precision (two indep. raters)
- Generate many candidate events (Freebase)

Event Selection becomes very important but challenging!

- 42 entities with 10k+ events
- 680k entities with 10+ events
2. Event Selection
Timeline Quality Criteria

1. Correctness

2. Relevance
Timeline Quality Criteria (cont.)

3. Content Diversity

Encourage selection of different entities

US Release | Award | EU Release

VS

Award | Award | US Release
4. Temporal Diversity / Layout

- Looks unaesthetic
- Hard to interact with events
Optimization Problem

Candidate Set:
1. Correctness (by construction)

Objective:
2. Relevance &
3. Content Diversity

Constraint:
4. Temporal Diversity

\[ T^* = \arg \max_{T \subseteq E} \text{Relevance}(T) \]
\[ \text{s.t. } \text{Temporal Diversity}(T) \]
(2) Relevance Signals

- Baseline: global relevance signal
- Based on # search queries for entity

- Biased towards popular but unspecific events
How to improve relevance signal

- Use **co-occurrences** on web scale

  "Robert Downey Jr" and "May 4, 2012" occurs 173 times on 71 different webpages

- US Release date of *The Avengers*
Assigning scores to entities/dates

- Run NLP tools across large web corpus of 10B documents (NER + CoRef)
- Extract entity-entity and entity-date co-occurrences within small windows
- **Normalize counts** using Normalized Pointwise Mutual Information (NPMI)
  - accounts for popular entities/dates
Improvements from New Signal

- Large improvement over global relevance
(3) Content Diversity

Encode diminishing returns in objective function:

Diminishing returns for including entity multiple times
(4) Temporal Diversity

- **Constraint** in optimization problem: “Event boxes cannot overlap.”
- **Enforce** **balanced layout** during optimization
Submodular Optimization

\[ T^* = \arg \max_{T \subseteq E} Relevance(T) \]

s.t. \( TemporalDiversity(T) \)

- Relevance(T) is submodular
  - “Diminishing returns”
  - Less reward for adding to “bigger set”

\[ X, Y \subseteq \Omega, \ X \subseteq Y, \ \text{and} \ x \in \Omega \setminus Y: \]
\[ f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y) \]
Algorithm & Theoretical Results

• Fast approximation through lazy-greedy
  ▪ Allows for zoom and interaction

• Provable approximation guarantee
  ▪ Worst-case: 33% of optimal solution!
  ▪ Still holds with complex constraint!
  ▪ We prove: constraint structure induces independence family that is a p-system
    (Calinescu et al. 2011)
3. Evaluation
Experimental Evaluation

• **User study** on Amazon Mechanical Turk
  - Rating of timelines challenging, possibly subjective, and no ground truth available

• **Pairwise comparisons** using 250 entities
  - Relative judgements
  - Explanations

• **Large-scale**
  - >1200 raters
  - >6000 tasks
Results

In what fraction of cases do raters prefer our full method over the ablated baselines?

- Baseline
- No Date Rel
- No Entity Rel
- No Temp. Div
- No Cont. Div

Global relevance signal
Removing Date/Entity Cooc
No content or temporal diversity

Full = “everything” = Global + Web Cooc + TD + CD
Related Work

- Document summarization: (Allan et al. 2001)
- Submodular optimization: (Krause & Golovin 2014), (Calinescu et al. 2011), (Nemhauser et al. 1978)
- Maps of information: (Shahaf et al. 2013)
- Timelines based on knowledge bases: (Mazeika et al. 2011), (Tuan et al. 2011), (Wang et al. 2010)
Conclusions

• **TimeMachine**: Automatic timeline generation for knowledge base entities

• **Submodular optimization framework**
  - Jointly optimizes for relevance, content diversity, and temporal diversity
  - Proved near-optimal performance guarantees

• **User studies** show
  - Web-based co-occurrence signals improve over baseline model (global importance)
  - Temporal and content diversity are crucial
Thanks!

Contact
@timalthoff  althoff@cs.stanford.edu

Check out the demo!
cs.stanford.edu/~althoff/timemachine

Paper / Proofs / Slides
cs.stanford.edu/~althoff

Acknowledgements
Evgeniy Gabrilovich, Arun Chaganty, Stefanie Jegelka, Karthik Raman, Sujith Ravi, Ravi Kumar, Jeff Tamer, Patri Friedman, Danila Sinopalnikov, Alexander Lyashuk, Jure Leskovec, David Hallac, Caroline Suen