The Future of Semantic Web Search

Oren Etzioni
Today a person is subjected to more new information in a day than a person in the middle ages in his entire life!
Well, we have Google...
Getting information off the Internet is like taking a drink from a fire hydrant.

Mitchell Kapor
Explosive growth of Mobile Devices
Getting information off the Internet is like taking a drink from a fire hydrant.

Mitchell Kapor
Imagine search over a semantic space

- Key words, docs $\rightarrow$ entities, predicates
- TF-IDF, pagerank $\rightarrow$ relation graph
- “10 blue links” $\rightarrow$ question answering

- Google Knowledge Graph (2012)
- Facebook Graph Search (2013)
Google’s split personality!
IE(text) = knowledge

I. Information Extraction (IE): what, how?
II. IE in action: reviews, news, tweets
III. Open IE = IE at Web scale (demo)
IV. Future Work

Technical Ideas: IE = ML

Scalability of IE
“By all accounts, Edison is credited as the inventor of the light bulb.”

invented(Edison, light bulb), 0.99
"You shall know a word by the company it keeps" (Firth, 1957)

- Nerja??
- …Nerja mayor…
- …Downtown Nerja…
- Spanish towns such as …Nerja…
- *Country* such as Garth Brooks…

Where do clues come from?
Target Relation = **Cities**
Training examples = **labeled** sentences

- Madrid **mayor**…
- …**Downtown** Madrid…
- Spanish cities **such as** …Madrid…
Basic Information Extraction Loop

1. Label sentences with examples of $R$
2. Extract frequent contexts of $R$ as **Clues**
3. Use **Clues** to find more examples

Options:

- Mutual recursion (Snowball)
- Negative examples (NELL)
- Parser
II. IE in Action---Product Reviews

- Hard to aggregate info across reviews
- Hard to filter phony reviews & outliers
- Hard to focus on key attributes
- Impossible to sort on attributes
- Difficult to read on phone

Pew Internet 2012 study:
24% used phone to look up reviews in store
1. La Carta de Oaxaca
Category: Mexican
Neighborhood: Ballard

**16**

2. El Camino
Categories: Mexican, Bars
Neighborhood: Fremont

**8**

3. The Saint
Categories: Mexican, Bars
Neighborhood: Fremont

**16**
Input: Seattle restaurant reviews + meta data

1. Extract key attributes = values
   A. Dim sum = delicious
   B. Server = rude

2. Cluster attributes into “food”, “service”...

Query: “good margaritas”

Answer: Sentiment sorted results
Extractive Interface on iphone

- Extracted Attributes
- Extracted Values
- Alternate Interfaces
- Green means Positive
- Mentions in Reviews
decide. No Regrets.
Eliminate Buyer’s Remorse
News in Consumer Electronics

Thousands of CE “news & rumor” sites, blogs, etc.
Wait for new model

Rumored: Next model coming out November 21

Verizon Samsung Galaxy Nexus Release Slated For Nov 21

MONDAY, 07 NOVEMBER 2011 22:27 STEVE TAYLOR

It looks like the Verizon Samsung Galaxy Nexus may actually be available online ahead of Black Friday 2011, something that many fans were hoping.

Samsung Galaxy Nexus release date leaked as November 21?

According to a Verizon roadmap uncovered by Droid-Life, the Samsung Galaxy Nexus release date online looks pegged for November 21. Now of course this doesn’t mean that you can’t grab it at a bargain price come Black Friday 2011 and Cyber Monday as we’re certain many retailers like Amazon and Wirefly would love to take advantage of the holiday shopping sale with one of the most hyped Android devices yet.
Some Current Examples

- https://www.decide.com/p/laptops-toshiba-p870-late-2012-satellite-17-3-106509622


- Even vacuum cleaners:

Critique of Conventional IE

• Extracting towns tells you nothing about hotels, countries, medications, etc.

• Hand-label training examples

**Difficult to scale!**
“No sentence left behind” IE

• Relations are not pre-specified
• No hand labeled examples
• Single pass over the corpus
• Independent of Genre (Spanish)
• Over 1,000 articles
First Web-scale, Open IE system
(Banko, Cafarella, Etzioni et al IJCAI ‘07)

1,000,000,000 distinct extractions

Peak Precision = 0.9 (limited recall)
Open source extractors:
reverb.cs.washington.edu
ollie.cs.washington.edu
Open IE Highlights

• Scale: 10B extractions
• No pre-determined vocabulary
• Entity linking & type inference
  – Graceful degradation
• Much work left…

Openie.cs.washington.edu
Sample of Extracted Relations

inhibits tumor growth in
is a person who studies
has a maximum speed of
gained fame as
was the first person to
has a PhD in
voted in favor of
died from complications of
granted political asylum to
identified the cause of
joined forces with
won an Oscar for
mastered the art of
is the patron saint of
wrote the book on
# Number of Relations

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yago</td>
<td>92</td>
</tr>
<tr>
<td>NELL</td>
<td>~500</td>
</tr>
<tr>
<td>DBpedia 3.2</td>
<td>940</td>
</tr>
<tr>
<td>PropBank</td>
<td>3,600</td>
</tr>
<tr>
<td>VerbNet</td>
<td>5,000</td>
</tr>
<tr>
<td>Wikipedia InfoBoxes, f &gt; 10</td>
<td>~5,000</td>
</tr>
<tr>
<td>TextRunner</td>
<td>100,000+</td>
</tr>
<tr>
<td>ReVerb</td>
<td>1,500,000+</td>
</tr>
</tbody>
</table>
## Event Calendar (2M tweets daily)

Alan Ritter (KDD ‘12), source code on github.

### January 2013

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>milk day: closed, inauguration, open</td>
<td>22</td>
<td>school: starts, open, closed</td>
<td>23</td>
<td>(today)</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>martin luther king day: school, closed, inaugurated</td>
<td></td>
<td>paramore: released, comes out, album</td>
<td>pebble: shipping, ship, talk</td>
<td>temple run 2: comes out, dropping, coming</td>
<td></td>
</tr>
<tr>
<td></td>
<td>blue monday: depressing, cheer, beat</td>
<td></td>
<td>itunes: be available, released, album</td>
<td>swansea: win, leg, game</td>
<td>glee: comes back, watch, excited</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obama: inauguration, inaugurated, sworn</td>
<td></td>
<td>hmv: accepting, vouchers, accepted</td>
<td>english: exam, class, study</td>
<td>english: exam, class, study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>inauguration: performing, perform, watch</td>
<td></td>
<td>english: exam, school, class</td>
<td>national pie day: celebrate, celebrating, heard</td>
<td>national pie day: celebrate, celebrating, heard</td>
<td></td>
</tr>
<tr>
<td></td>
<td>more...</td>
<td></td>
<td>more...</td>
<td>more...</td>
<td>more...</td>
<td>more...</td>
</tr>
<tr>
<td>28</td>
<td>hard rock: compete, beginning, entered</td>
<td>29</td>
<td>justin: album, listen, released</td>
<td>30</td>
<td>rim: launch, sends out, unveil</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>born sinner: album, released, drops</td>
<td></td>
<td>oppo find 5: date, retail, arrive</td>
<td>blackberry: launch, arrive, launched</td>
<td>microsoft: users, offering, ending</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cole: album, drops, born</td>
<td></td>
<td>destiny: album, release, releasing</td>
<td>irs: accepting, start, filing season</td>
<td>windows phone 7.8: update, start, aiming</td>
<td></td>
</tr>
<tr>
<td></td>
<td>more...</td>
<td></td>
<td>more...</td>
<td>more...</td>
<td>more...</td>
<td>more...</td>
</tr>
<tr>
<td>26</td>
<td>australia day: party, celebrate, celebrating</td>
<td>27</td>
<td>royal rumble: win, watch, match</td>
<td>fa cup: match, trafford, game</td>
<td>pro bowl: watch, played, game</td>
<td></td>
</tr>
<tr>
<td></td>
<td>holocaust memorial day: find out, concert, programme</td>
<td>28</td>
<td>coco: concert, show, meet</td>
<td>miami: concert, show, meet</td>
<td>more...</td>
<td></td>
</tr>
</tbody>
</table>
IV. Future Work

• Open Question Answering
  – NL QA over arbitrary text (Fader, ACL ‘13)
  – Robust dialog “..ok, let’s look for Sushi outside of downtown, but I need convenient parking..”

• Inference
  – Often information is implicit
  – Text meets Linked Open Data
Comprehensive Knowledge

Open-ended Learning

Tractable reasoning
Machine Reading

Auto-Text to Knowledge
Conclusions

• Keyword search ➔ extraction, relations
• IE is here: Decide.com, revminer.com
• IE = ML
  – Where does labeled data come from?
• **Open IE:** leverage the structure of language to achieve scale
• **Open Question Answering & reasoning**
  – Thank you!
Example Opinion Patterns

- X is so loving
- I love X so much
- X makes me laugh
- I am in love with X
- X did a great job
- It was a pleasure working with X
- I’d be lost without X
- I hate X with a passion
- X is an idiot
- Death to X
- X is gross
- X is stupid as hell
- I don’t like X one bit
- I loathe X
- X is my mortal enemy
What prevents X? =

- What can people do to prevent X?
- How can X be avoided?
- What will help prevent X?
- What are the steps to reduce X?
- How can you prevent X?
- What is the prevention of X?
- Is there a way to prevent X?
IE Systems Discover Clues

tears of_
feelings such as_
heart filled with_
heart was filled with_
heart is filled with_
heart was full of_
feelings, such as_
string of_
pang of_
emotion such as_
heart is full of_
intense feelings of_
overwhelming
feelings of_

deep feelings of_
mixed feelings of_
I was overcome with_
emotions, from_
feelings of intense_
strong feelings of_
I am filled with_
hearts filled with_
feelings of deep_
feelings of extreme_
paroxysms of_
I'm filled with_
source of deep_
he was filled with_

← Earliest
Most recent →

I just burned.
People feel a big vote.
I have been.

world looks
other country.
expressed.
I was fell.
issue is danger.
matters of.
sheer drill.
Majesty.

Association.
expressed.
browser.