Learning Causation from Text

Jaime Arguello
Jon Elsas
Thuy Linh Nguyen
Examples from Google

- Phrases that indicate causation
  - A “caused” B
  - B “is caused by” A
  - A “leads to” B
- Mining Google (or PubMed / Medline) for causal relationships
- New patterns & event pairs added to our working set as system iterates
Examples from Google (2)

"Air rage is caused by oxygen deprivation."--Diana Fairechild
An air travel guide about air rage, recycled air, lost luggage, limited carry-ons, jetlag, flight attendants, breathing, environment, recycled air, ...
www.flyana.com/rage.html - 35k - Cached - Similar pages

I stand corrected: AIDS is caused by a mycoplasma made in the USA
Returning to Dirk Chardet, he asserts that "AIDS is caused by a mycoplasma (a tiny cell without walls) that has a visna lentivirus insertion (this virus ...
www.onlinejournal.com/artman/publish/article_115.shtml - 32k - Cached - Similar pages

bmj.com Rapid Responses for Sidley, 324 (7335) 446
bmj.bmjournals.com/cgi/eletters/324/7335/446/a - Similar pages

Oral Herpes is Caused by Herpes Simplex Viruses 1 and 2- iVillage
Oral herpes does not usually feature symptoms, however cold sores or blisters on the lips and mouth may occur. In some cases, people with oral herpes never ...
health.ivillage.com/dental/dmouthsores/topics/0,,4vfl,00.html - 47k - Cached - Similar pages
Examples from Google (3)

"Air rage is caused by oxygen deprivation." -- Diana Fairechild
An air travel guide about air rage, recycled air, lost luggage, limited carry-ons, jetlag, flight attendants, breathing, environment, recycled air, ...
www.flyana.com/rage.html - 35k - Cached - Similar pages

"Airplane air spreads viruses." -- Diana Fairechild
Oxygen Deprivation (Hypoxia) Emergency Oxygen · Air Rage · Sleeping & Poor Air · Preemies & Poor Air · Airline Bankruptcies & Poor Air ...
www.flyana.com/air.html - 30k - Cached - Similar pages
[More results from www.flyana.com]

Peace Talks
(I have a) theory that air rage is triggered by oxygen deprivation on the airplane.
The inadequate oxygen to the brain fuels passengers’ temper tantrums ...
www.goodradioshows.org/peaceTalksL15.html - 18k - Cached - Similar pages

Air Rage Caused by Oxygen Deprivation
Description, This piece on "air rage" was first written for and published by ...
www.uncoverthenet.com/listing/6911.php - 10k - Cached - Similar pages
Bad Examples from Google!

Connecticut cause viagra blindness
The FDA is investigating whether Viagra causes blindness. Maybe there's something to ... Or maybe the FDA has it backwards: maybe blindness causes Viagra. ...
113451.q.0wc.info/ - Similar pages

Amazon.com: Critical Mass: How One Thing Leads to Another: Books ...
Amazon.com: Critical Mass: How One Thing Leads to Another: Books: Philip Ball by Philip Ball.
www.amazon.com/exec/obidos/tp/detail/-/0374281254?v=glance - 100k - Cached - Similar pages

Cryptome
It is a result of 15 researcher years, hundreds of interviews, and in-depth analysis of more than 5000 documents. We at Homeland Security Research Corp. are ...
cryptome.org/ - 86k - Jan 21, 2006 - Cached - Similar pages
Our proposal

• Seed the system with positive and negative examples of both contexts (+/-s_i) and causal relationships (+/-(A,B))
  – +s_1 = “is a cause of”, +s_2 = “directly leads to”,
    -s_1 = “does not cause”, -s_2 = “is not related to”
  – + (smoking, cancer), + (greenhouse gasses, global warming), - (MMR, autism), - (…) 
  – Initial probabilities:
    P( +c | “is a cause of”) = 1 
    P( +c | “does not cause”) = 0 
    P( +c | (smoking, cancer)) = 1 …
Our proposal

• Run a query for each context, $s_i$, and for each result returned $(A_i, B_i)$ compute the probability that $(A_i, B_i)$ is a causal relationship:
  
  $- P( +c \mid (A_i, B_i)) = \sum_k P( +c \mid s_k) P( s_k \mid (A_i, B_i))$

• Do the same for retrieving contexts given entity-pairs:
  
  $- P( +c \mid s_i) = \sum_k P( +c \mid (A_k, B_k)) P( (A_k, B_k) \mid s_i)$
Our proposal

• Tentative Evaluation Strategy
  – 10 domain experts label 100 causal relationships as binary plausible/not_plausible
  – 333 (2/3) majority-vote judgments
  – Precision/Recall/F-measure

• Additional Resources from Unified Medical Language System (synonyms/hyponyms)
  – Metathesaurus
  – Semantic Network
  – Specialist Lexicon
Potential Heuristics Improvement

- World knowledge tells us that there exist sets $\text{SET}_a$ and $\text{SET}_b$ for which $P(x \rightarrow y \mid x \in \text{SET}_a \& y \in \text{SET}_b)$ is particularly high.

- Examples:
  - drugs $\rightarrow$ side effects
  - forms of crime $\rightarrow$ forms of punishment
  - forms of altruism $\rightarrow$ forms of reward

- Thus, our bootstrapping approach can be:
  - We know drugs have side effects, and we’re looking for “contexts that mark causality”.
  - (1) run query “is a drug”, we find “x is a drug”.
  - (2) run query “x causes”, we find “x causes y”
  - (3) run query “y is a”
  - (4) if y is a side-effect, then we can be more confident of those patterns where x and y occur close to each other, since $x \in \text{drug}$ and $y \in \text{side-effect}$