이벤트 탐지 연구 동향 및 응용 사례

정유철
enthusia77@gmail.com

Outline

- Introduction & Backgrounds
- What is Event?
- Examples of Events
- Event Extraction
- Challenging Research Issues
- Possible Applications
- Concluding Remarks

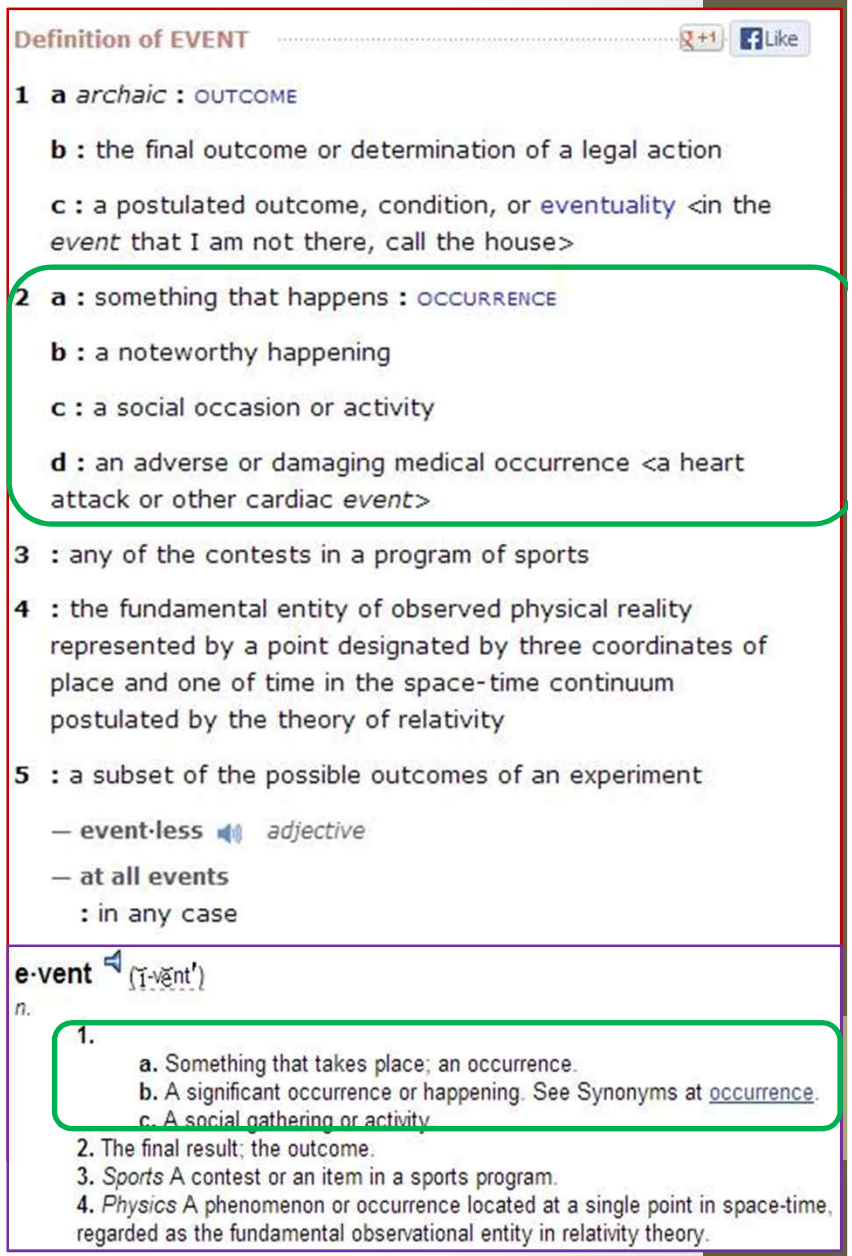
Introduction

- Long history of the needs to identify human activities (or events) surrounding user behaviors
- Increasing number of user-generated contents (e.g., social media data)
- Several event extraction approaches in terms of text mining
- Interesting applications concerning daily lives of human beings

What is Event?

- Definition of Event
  - 사람의 일상에서 일어나는 행위
  - 일어나는 (중요한) 무엇; 사건
  - 사회적 행사, 이벤트
    (문화, 스포츠, 등)

+ 기타 인간의 activity들과 관련된 행위



Definition of EVENT

1 **a** *archaic* : OUTCOME

b : the final outcome or determination of a legal action

c : a postulated outcome, condition, or eventuality <in the *event* that I am not there, call the house>

2 **a** : something that happens : OCCURRENCE

b : a noteworthy happening

c : a social occasion or activity

d : an adverse or damaging medical occurrence <a heart attack or other cardiac *event*>

3 : any of the contests in a program of sports

4 : the fundamental entity of observed physical reality represented by a point designated by three coordinates of place and one of time in the space-time continuum postulated by the theory of relativity

5 : a subset of the possible outcomes of an experiment

— **event·less** *adjective*

— **at all events**
: in any case

e·vent (ĭ-vĕnt′)
n.

1.
a. Something that takes place; an occurrence.
b. A significant occurrence or happening. See Synonyms at occurrence.
c. A social gathering or activity.

2. The final result; the outcome.
3. *Sports* A contest or an item in a sports program.
4. *Physics* A phenomenon or occurrence located at a single point in space-time, regarded as the fundamental observational entity in relativity theory.

Backgrounds

- Lots of previous work on Newswire (Event Extraction)
  - MUC scenario template 1987 – 1998
    - MUC-3/4: terrorist incidents
    - MUC-6: executive succession
  - ACE 2005
    - Eight types of events: Life (Be-Born, Marry, Divorce, Injure, Die), Movement (Transport), Transaction (Transfer-Ownership, Transfer-Money), Business (Start-Org, Merge-Org, Declare-Bankruptcy, EndOrg), Conflict (Attack, Demonstrate), Contact (Meet, Phone-Write), Personnel (Start-Position, End-Position, Nominate, Elect), Justice (ArrestJail, Release-Parole, Trial-Hearing, Charge-Indict, Sue, Convict, Sentence, Fine, Execute, Extradite, Acquit, Appeal, Pardon)
  - Timebank (detecting temporal association)
    - Seven types of events: reporting, perception, aspectual, i_action, i_state, state, and occurrence

Drawbacks

-Limited to narrow domains

-Performance is still not great

Backgrounds

- A few related works on social media data [Story, Event]
  - Blogs
    - Capturing story events for storytelling [A. S. Gordon, 2010]

#Drawbacks
- Need to identify the implicit discourse structure of stories

  - Twitter
    - Mundane events extraction from Twitter / Event type classification [A. Ritter, et al., 2012]

#Drawbacks
- Hard to segment entity and entity phrases
- Redundancy of information

Backgrounds

- Event Mining in Multimedia Streams

5W1H of Multimedia Events

Actions, activities, and their aggregates

Example of a real-world event : Game 4 of 2007 NBA Finals

[Excerpted from L. Xie, et al. 2008]

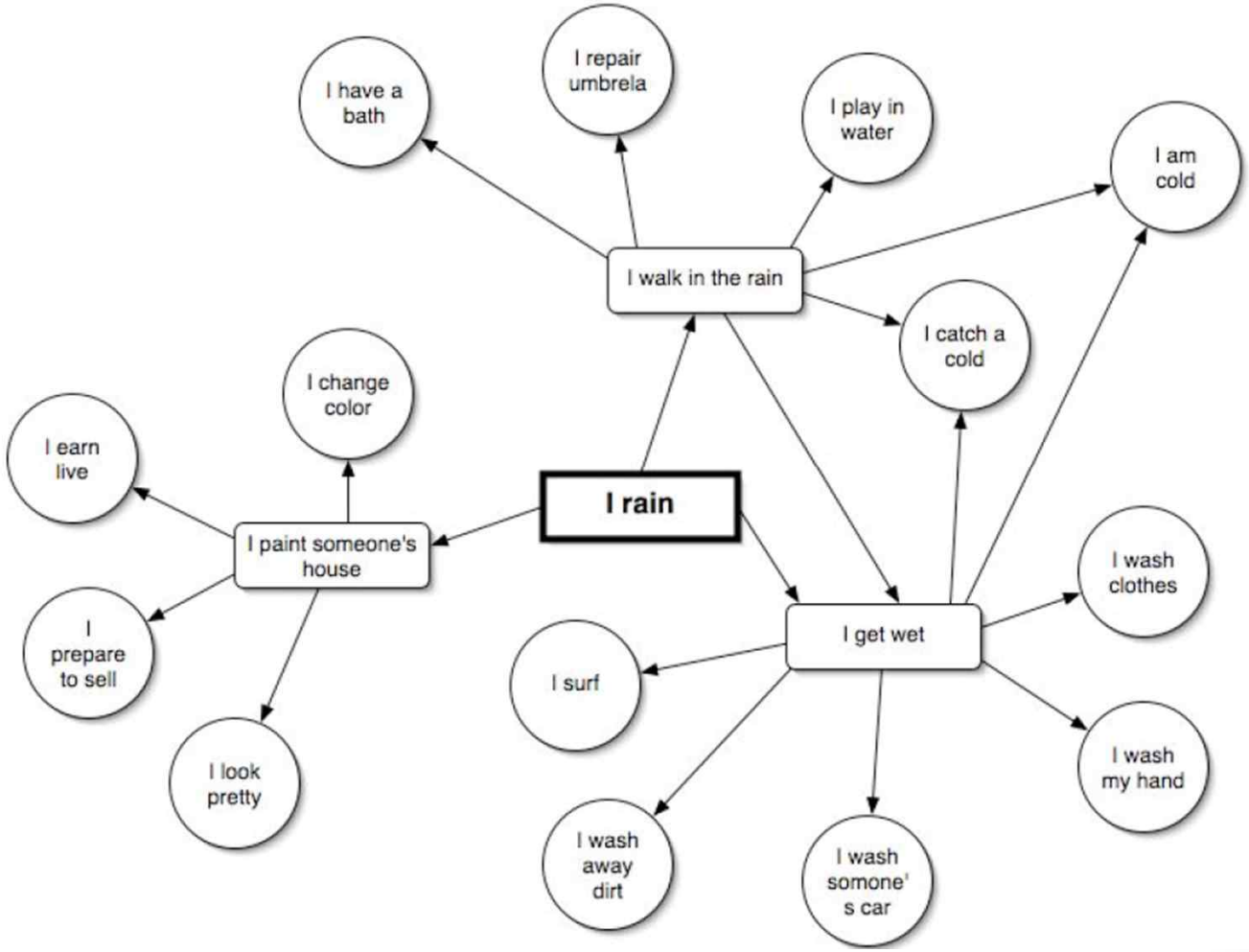
Examples of Events (1/3)

- Events in Commonsense
  - eating & drinking, running, explosions, questioning, searching, shouting, public speaking, automobile driving, chasing, conversation, cleaning, falling, fires, hiding, holidays, leisure, shooting, sleeping, walking, war, cheering, death, discussion, hand clapping, music, raising hands, reading, swimming, examinations, measuring, painting, shopping, smoking, starvation, writing, experiments, jumping, lifting & carrying, navigation, children playing, fighting, marriage, parties, singing, travel, animal feeding, debates, digging, ocean travel, farming, forecasting, gambling, etc.

[A. S. Gordon, 1999]

Examples of Events (2/3)

- EventNet from Open Mind Commonsense Knowledge



[J. Espinosa and H. Lieberman, 2005]

Examples of Events (3/3)

- Events from Twitter Data

Steve Jobs **died** Oct. 6
iPhone 5 **announcement** coming Oct 4th
..

Entity	Event Phrase	Date	Type
Steve Jobs	died	10/6/11	DEATH
iPhone	announcement	10/4/11	PRODUCTLAUNCH
GOP	debate	9/7/11	POLITICALEVENT
Amanda Knox	verdict	10/3/11	TRIAL

Event Extraction

- Rule-based extraction
  - Mostly, lexico-syntactic and lexico-semantic [J. Borsje, et al., 2011]
- Supervised learning (e.g., CRFs) based extraction
  - Various features including lexical, semantic (entity type, role ..) , dependency, syntactic, relative-position, etc. [L. Hou, et al., 2012][A. Ritter, 2012]
- Semi-supervised learning (i.e., Bootstrapping) based extraction
  - Automatically learn how to recognize events involving multiple entities and their roles [T. Liu and T. Strzalkowski, 2012]

Event extraction from Twitter (1/4)

- Step1: Named Entity Recognition (NER) for Twitter
  - Sequence Labeling Task
    - IOB encoding
  - Conditional Random Fields
  - Features:
    - Orthographic:
      - Prefixes, suffixes
    - Dictionaries:
      - type list gathered from Freebase
      - Brown clusters
      - Outputs of T-POS, T-CHUNK, and T-CAP
    - Contextual
      - POS tags
      - Adjacent words

Word	Label
T-Mobile	B-ENTITY
to	O
release	O
Dell	B-ENTITY
Streak	I-ENTITY
7	I-ENTITY
on	O
Feb	O
2nd	O

[A. Ritter, 2011]

Event extraction from Twitter (2/4)

- Event referring phrases
  - Useful to display in connection with events
    - E.g. "**Steve Jobs**" + [ ] + "**October 6**"

entity

date

Examples:

Apple to **Announce** iPhone 5 on October 4th! YES!

iPhone 5 **announcement** coming Oct 4th

WOOOHOO **NEW** IPHONE TODAY! CAN'T WAIT!

[A. Ritter, 2012]

Event extraction from Twitter (3/4)

- Training Data Annotation
  - Annotated 1,000 tweets (19,484 tokens)
  - Similar to EVENT tags in TimeBank
  - Sequence-labeling problem
    - IOB encoding
    - Employs Conditional Random Fields (CRFs)

[A. Ritter, 2012]

Event extraction from Twitter (4/4)

- Step 2: Event Phrase Extraction [A. Ritter, et al., 2012]
  - Sequence Labeling Task
    - IOB encoding
  - Conditional Random Fields
  - Features:
    - Contextual features
      - POS tags
      - Adjacent words
    - Dictionary Features
      - Event words gathered from WordNet
      - Brown Clusters
    - Orthographic Features
      - Prefixes, suffixes

[A. Ritter, 2012]

Discussions

- Limitations of event extraction
  - Supervised EE training is very expensive …
    - Lots of types of events
    - Lots of paraphrases of each event
    - Event annotation is slow (because information is complex)
  - So semi-supervised methods are particularly attractive
    - Some research shows that its performance outperforms that of supervised EE.
    - However, open domain event extraction requires a lot of guidance according to event types (or cases)

Challenging Research Issues

- Event Type Classification

  → Analyze topicality of the given event phrase

- Semantic Matching of Heterogeneous Events

  → Dealing with unknown events by employing knowledge resources

- Web Mining for Event-Based Commonsense Knowledge

  → Automatically generating event-based commonsense knowledge

- Activity Knowledge Base Construction

  → Enriching activity/event related knowledge for activity-based applications

Event Type Classification (1/2)

 # Situation

- Many Different Types
- Not sure what is the right set of types
- Set of types might change
  - Might start talking about different things
  - Might want to focus on different groups of users

- Unsupervised event type induction
  - Latent variable models (e.g., LDA)
    - Generative probabilistic models
  - Categorize events into types, for example:
    - Sports, concert, perform, movie, product, school, religion, prize, etc.

- Advantages:
  - Discovers types which match the data
  - No need to annotate individual events
  - Don't need to commit to a specific set of types
  - Modular, can integrate into various applications

[A. Ritter, et al., 2012]

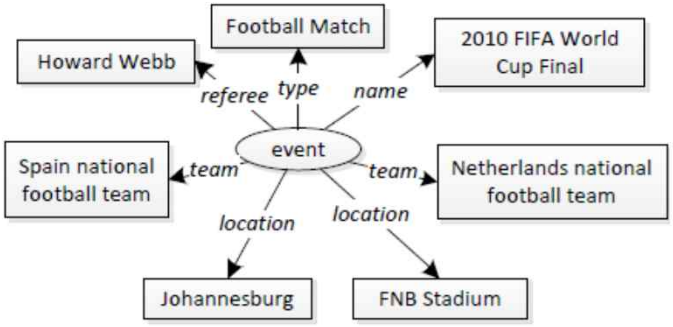
Event Type Classification (1/2)

Label	Top 5 Event Phrases	Top 5 Entities
Sports	tailgate - scrimmage - tailgating - homecoming - regular season	espn - ncaa - tigers - eagles - varsity
Concert	concert - presale - performs - concerts - tickets	taylor swift - toronto - britney spears - rihanna - rock
Perform	matinee - musical - priscilla - seeing - wicked	shrek - les mis - lee evans - wicked - broadway
TV	new season - season finale - finished season - episodes - new episode	jersey shore - true blood - glee - dvr - hbo
Movie	watch love - dialogue theme - inception - hall pass - movie	netflix - black swan - insidious - tron - scott pilgrim
Sports	inning - innings - pitched - homered - homer	mlb - red sox - yankees - twins - dl
Politics	presidential debate - osama - presidential candidate - republican debate - debate performance	obama - president obama - gop - cnn - america

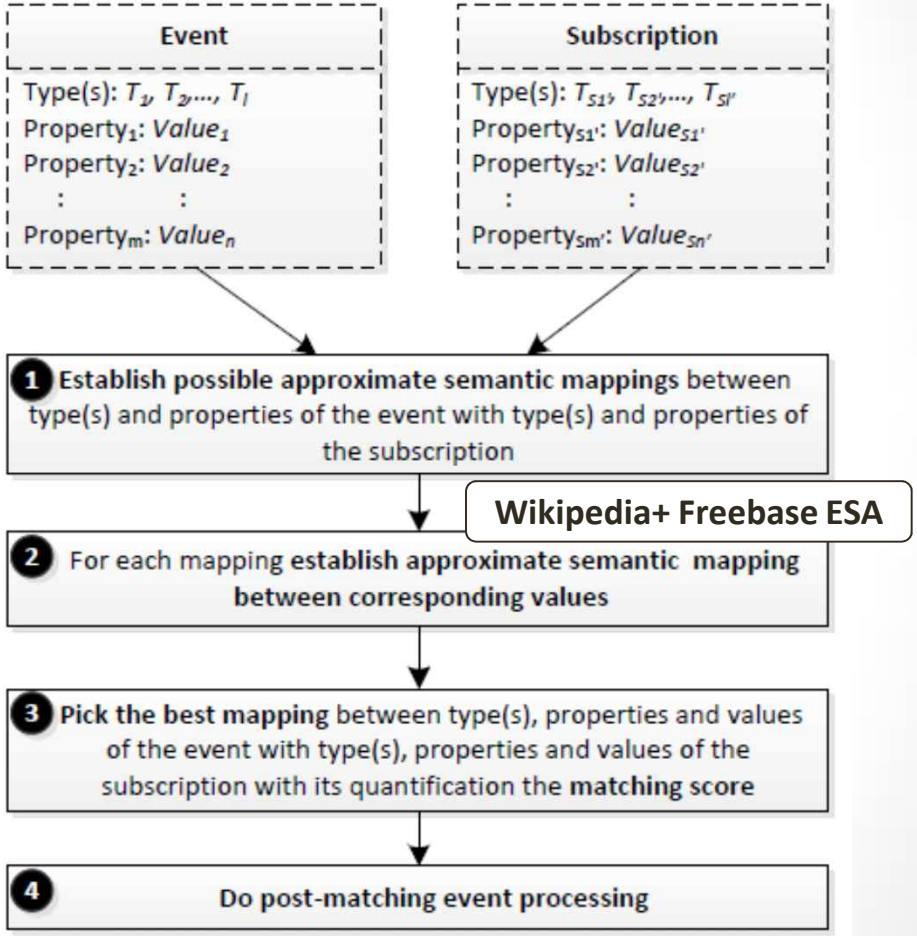
Event type discovered by LDA model

[A. Ritter, et al., 2012]

Semantic Matching of Heterogeneous Events



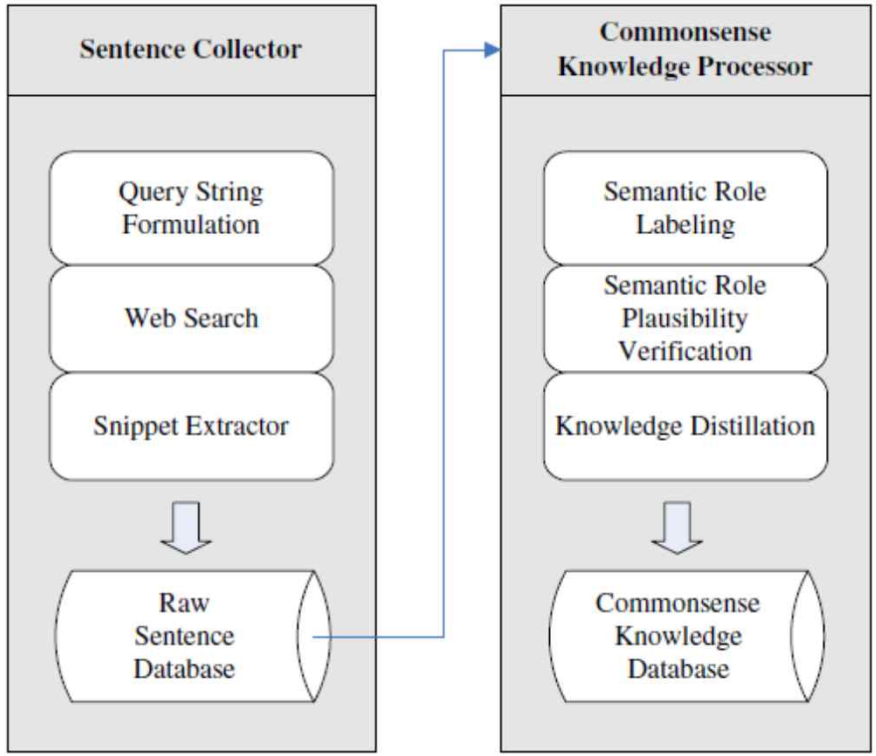
An example event



Event matching procedures

Excerpted from [S. Hasan, et al., 2012]

Web Mining for Event-Based Commonsense Knowledge



A framework for knowledge extraction process

Lexico-syntactic patterns for "capability" of the subject

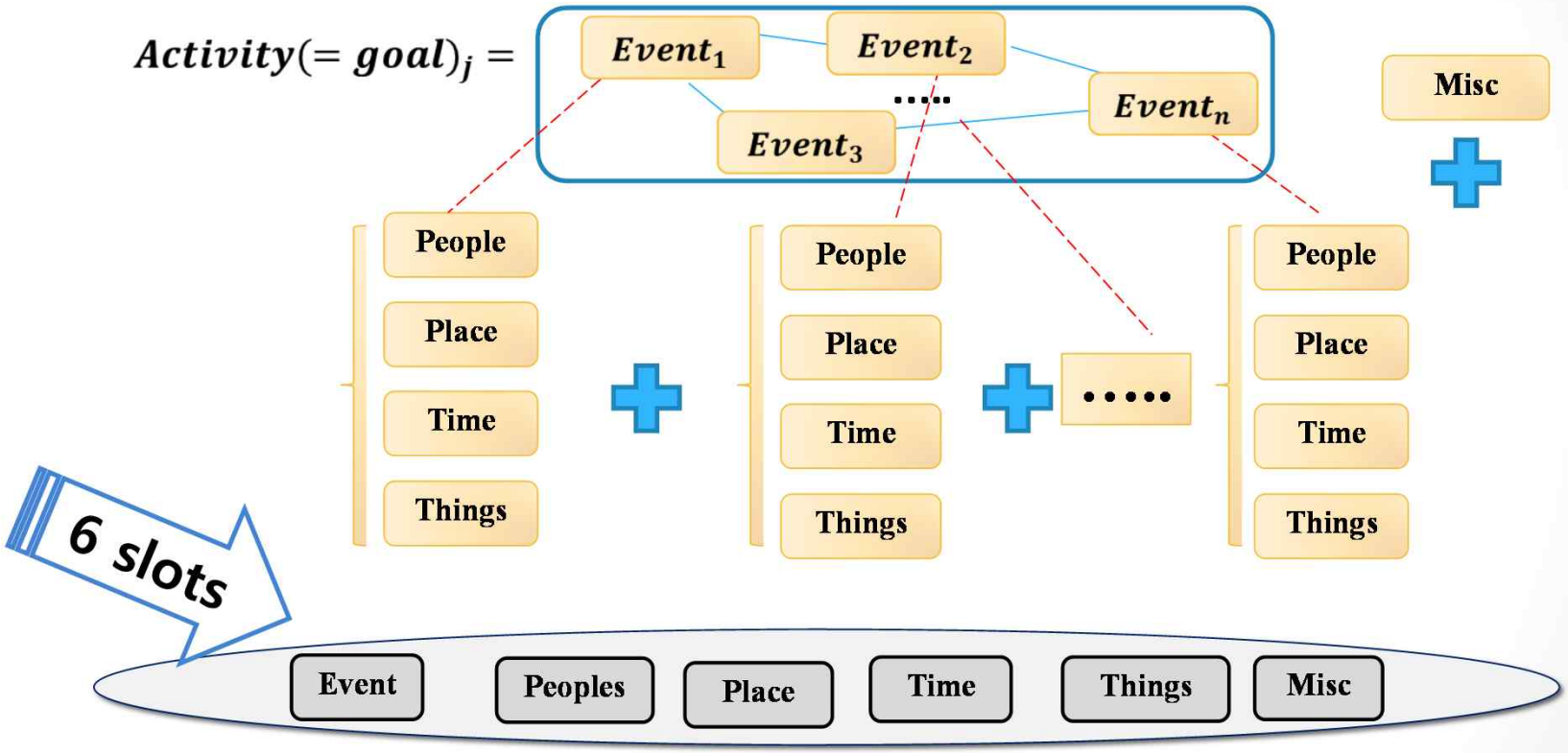
(1) subject + *"is capable of"* + *verb* (present participle) (e.g., "A soldier is capable of fighting the enemy")
(2) subject + *"is able to"* + *verb* (e.g., "The soldier is able to fire the gun")
(3) subject + *modal verbs* (can, will, should) + *verb* – (e.g., "A soldier can fight the enemy.")

ARG0	Verb	ARG1
Dog	Eat	Foods
Dog	Bite	Someone
Fish	Enter	The pool
Police	Check	The taxi's licenses
Police	Capture	A criminal
Person	Hack	The system
Person	Enjoy	Hiking and wildlife watching
Soldiers	Stop	The violence
Soldier	Obey	The army

Commonsense knowledge retrieved

Excerpted from [S.-H. Hung, et al., 2010]

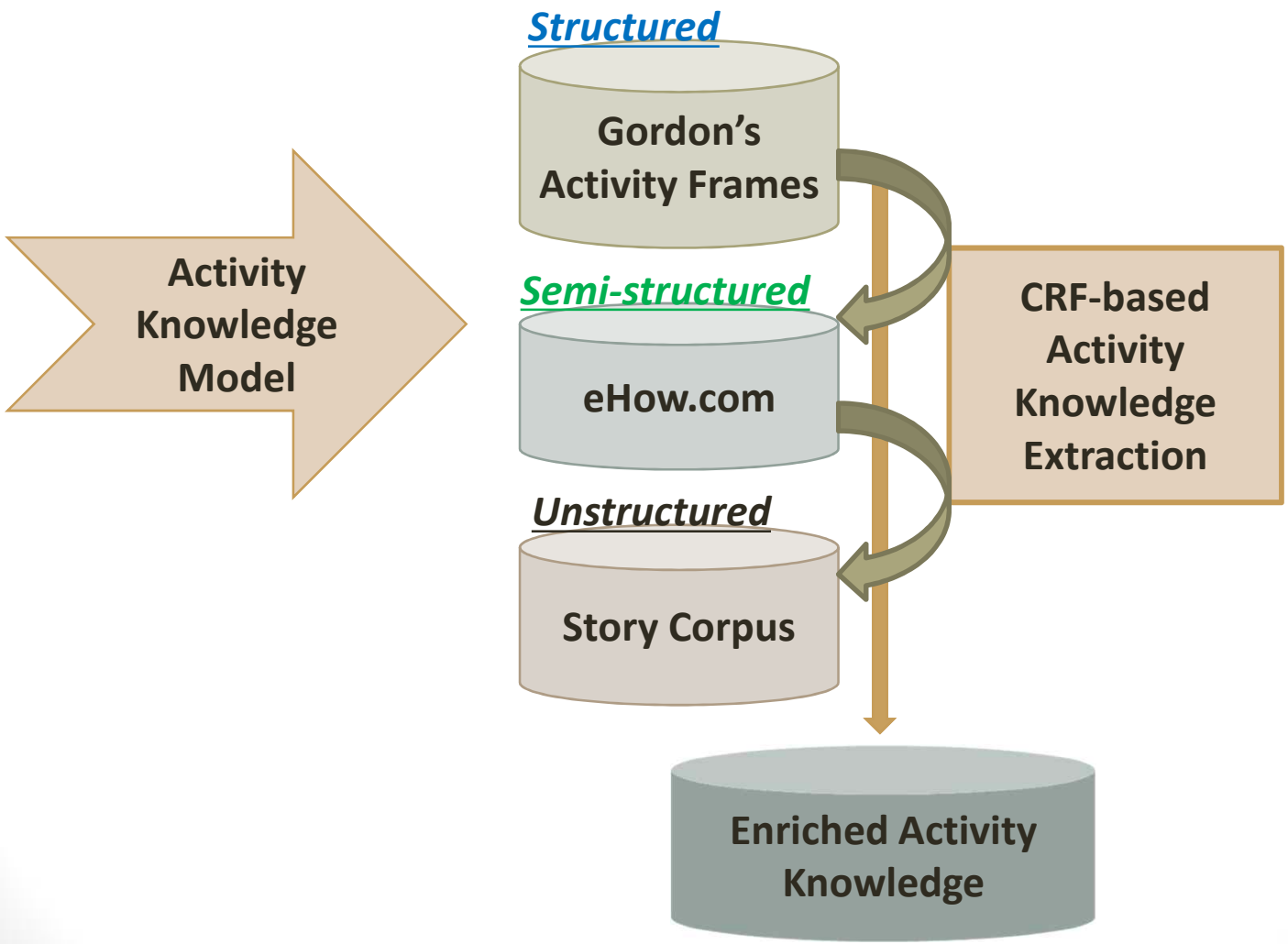
Activity Knowledge Base Construction (1/2)



A Knowledge Representation Model

Excerpted from [Y. C. Jung, 2010]

Activity Knowledge Base Construction (2/2)



Excerpted from [Y. C. Jung, 2010]

of Slot Values in Enriched Activity KB

35 Activities	# place	# time	# people	# event	# thing	# misc
a wedding reception and party	60	18	144	172	332	32
a woman giving birth	12	8	55	40	9	27
applying and interviewing for a job	7	27	108	113	88	23
asking someone out on a date	6	8	0	53	31	22
cleaning a house	25	16	1	148	75	5
doing the laundry at home	4	45	10	167	416	6
donating blood at a blood drive	64	33	34	89	89	7
exploring a cave	35	10	12	38	66	19
finding and checking out a book from a library	30	3	12	39	95	13
going gambling in a casino	90	3	34	150	170	37
going skiing in the mountains	146	9	6	107	80	2
going surfing in the ocean	28	4	31	100	73	73
going to a dog show	39	23	60	179	212	36
going to a farmer's market	40	39	68	140	268	52
going to a fast food restaurant	45	6	6	33	326	48
going to a flea market	84	36	73	176	273	39
going to a horse show	56	32	38	110	241	36
going to bed at night	33	43	33	60	28	33
going to the race track to bet on horse races	44	6	20	202	223	46
going to the zoo	228	16	39	51	188	4
going trick-or-treating on halloween	80	17	119	77	170	37
going water skiing on a lake	19	3	30	51	95	7
having a drink in a bar	38	8	61	58	146	10
having a garage sale	9	29	52	147	232	17
having a tea party	18	14	56	74	366	13
living in a nursing home	52	14	71	45	63	42
playing in a marching band	3	0	89	66	120	33
reading the newspaper	0	5	13	47	148	17
riding on a roller coaster	24	3	20	49	67	41
running in a marathon	50	119	69	330	144	102
shopping in a grocery store	83	8	16	40	243	0
shoveling snow in the winter	34	11	5	115	177	2
using a computer for internet access	4	1	0	42	166	7
waking up in the morning	22	36	0	58	69	0
writing a poem	4	8	27	135	245	75
Total (permit duplications)	1516	661	1412	3501	5734	963
Total (distinct instances)	788	522	651	1851	2907	671

Gordon's 768

3,676 → 7,390

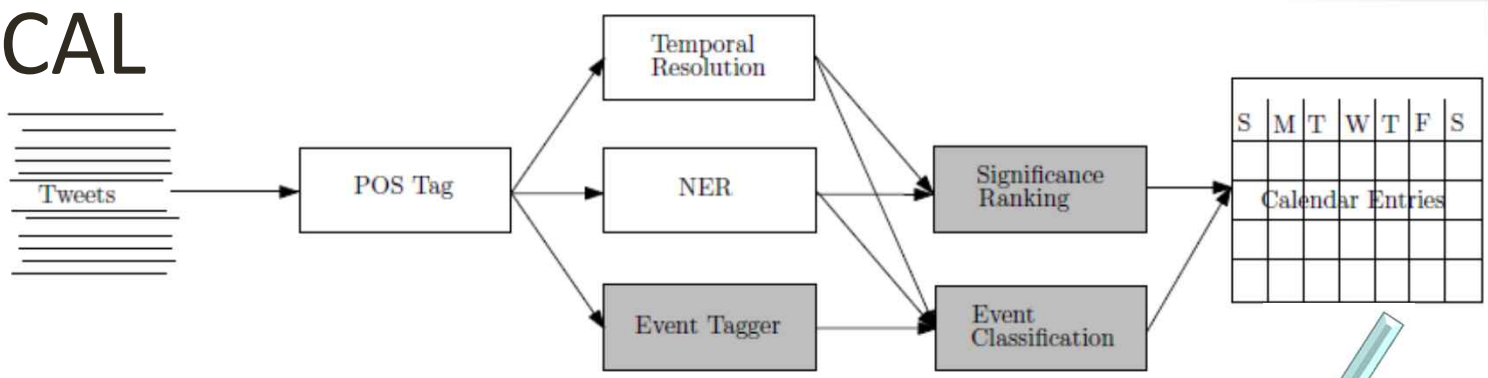
Excerpted from [Y. C. Jung, 2010]

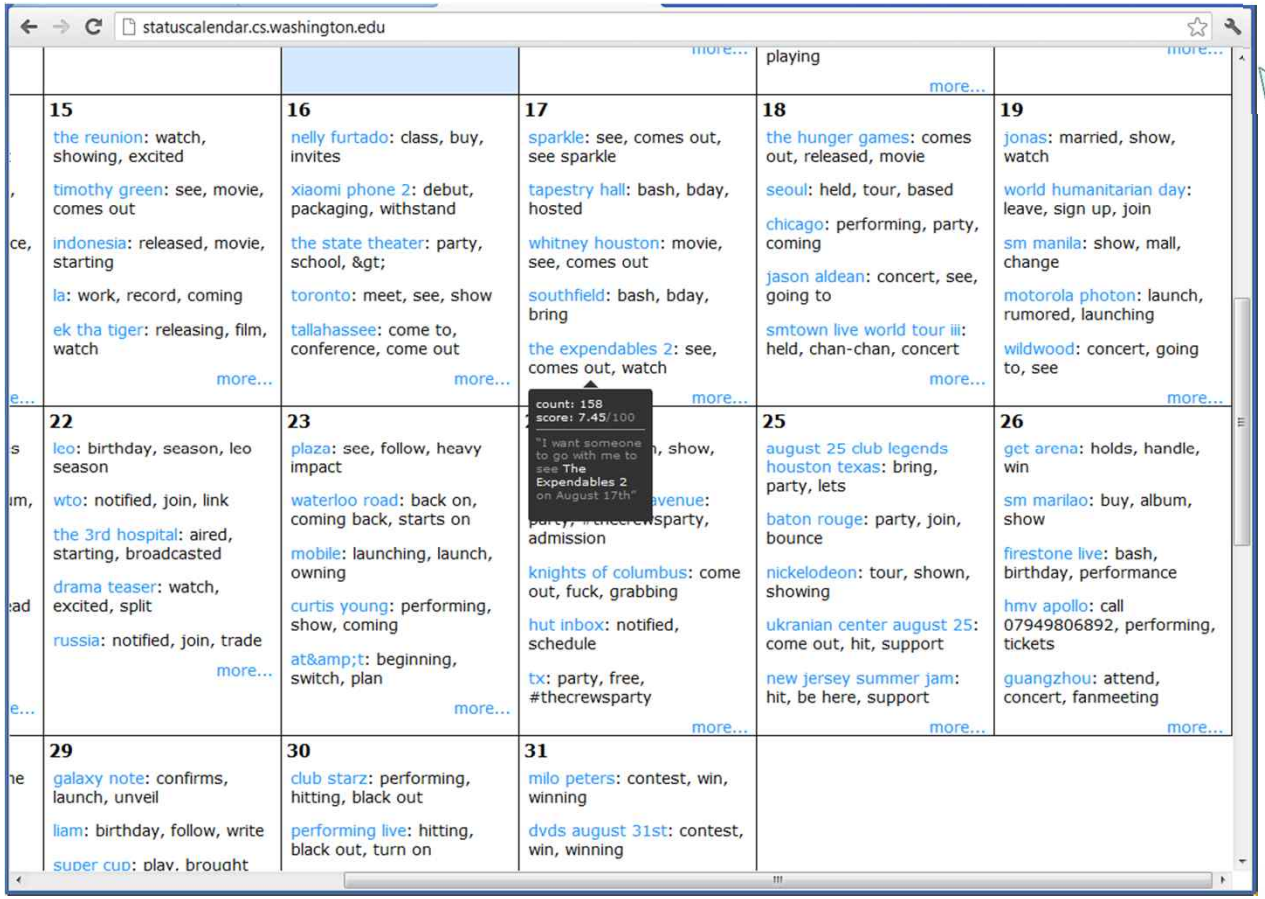
Possible Applications

- **Calendar and Activity Prediction**
  - TwiCAL [A. Ritter, 2012]
  - Activity Prediction [W. Weerkamp, 2012]
- **Event-Driven Search**
  - Event-centric search [J. Strotgen and M. Gertz, 2012]
- **Various Types of Context-aware Applications**
  - Query-free search
  - Situation prediction / Intent analysis

TwiCAL

Pipeline for extracting events from Twitter





Excerpted from [A. Ritter, 2012]

Activity Prediction

Table 1: Tweets referring to past or current activities.

Frustrated with a few things this morning. Most of all, that I set the wrong alarm and slept through run time... GRR
I did not sleep at all last night. I must be that excited for Peter and Chris tonight.
Working on my day off....I wouldn't have it any other way!!!
Bird watching with momma
just finished watching Wrath Of The Titans! #GoodMovie :)

'Life mining'
= "extracting useful knowledge from the combined digital trails left behind by people who live a comfortable part of their life on-line"

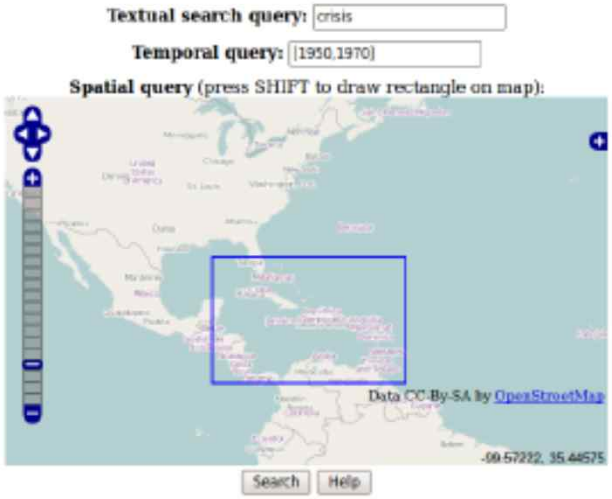
"Activity prediction is a special case of life mining"

Table 2: Tweets referring to a future activity.

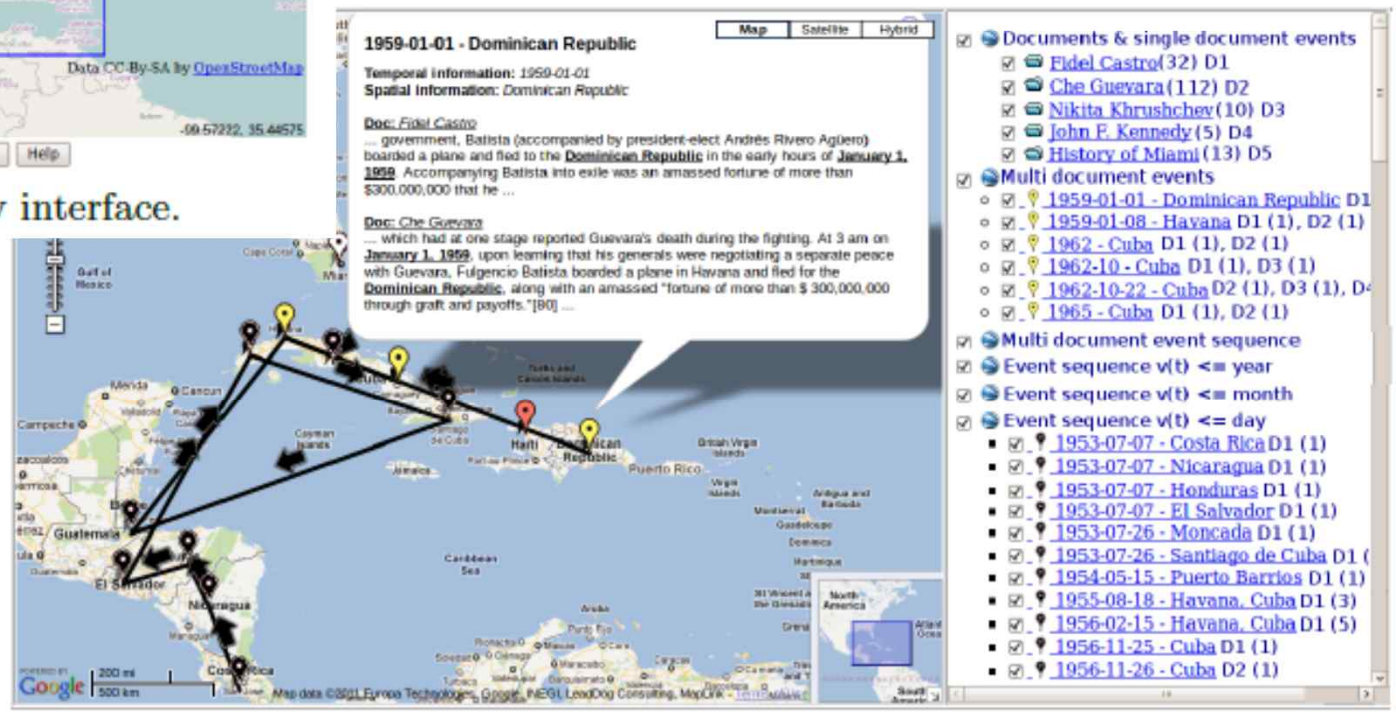
im gonna wrestle a midget tonight... uberexcited
I'm kinda nervous for my date tonight!! #NeedToManUp
exited to dance with the girls tonight:) #wewillrockyou
Excited for bodypump class tonight! :D #gym #motivated #excited
come on germany tonight.. like to see u in the final.... watching tv tonight at home... have the dogs.. kids out public viewing

Excerpted from [W. Weerkamp, 2012]

Event-centric search



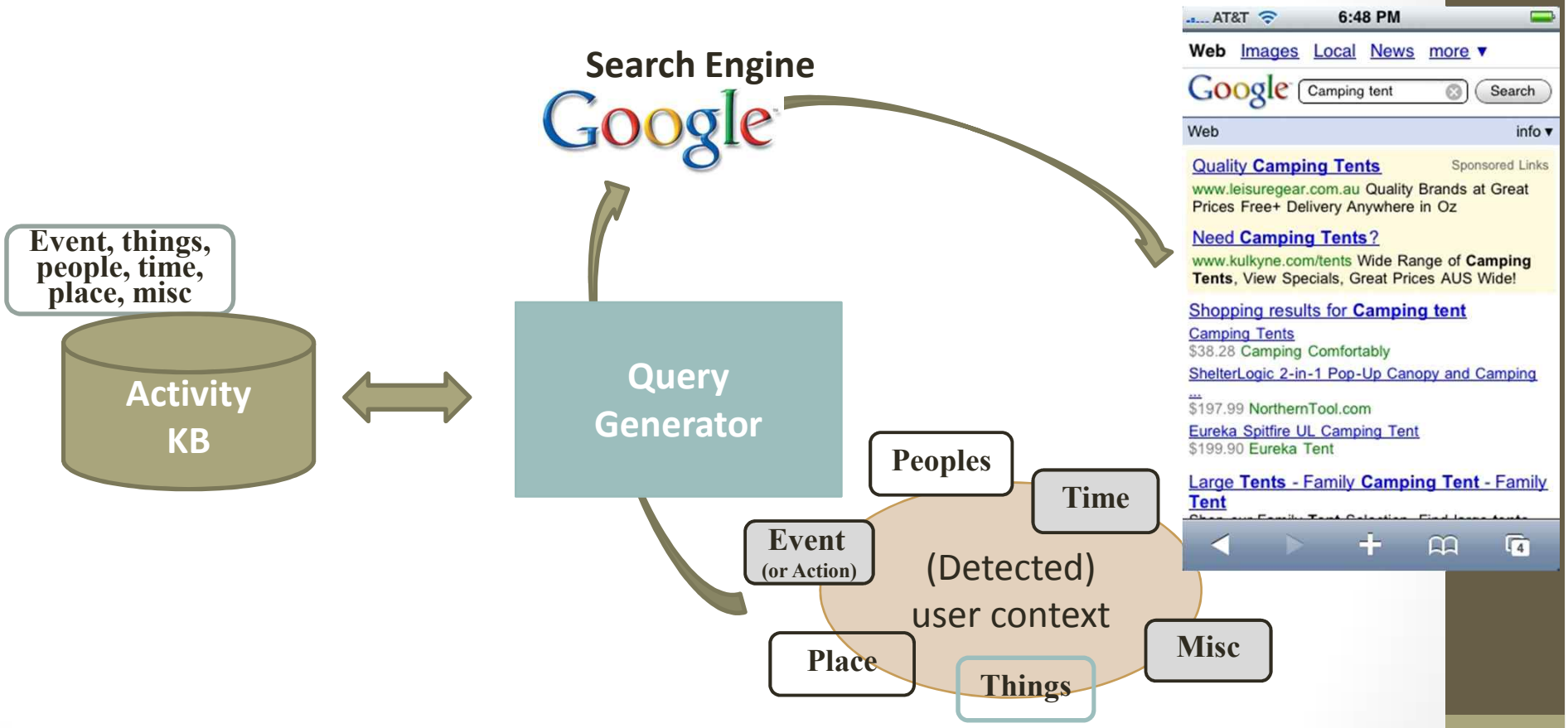
(a) Query interface.



(b) Map-based visualization of event sequence with v(t) $\leq$ day for query in (a).

Excerpted from [J. Strotgen and M. Gertz, 2012]

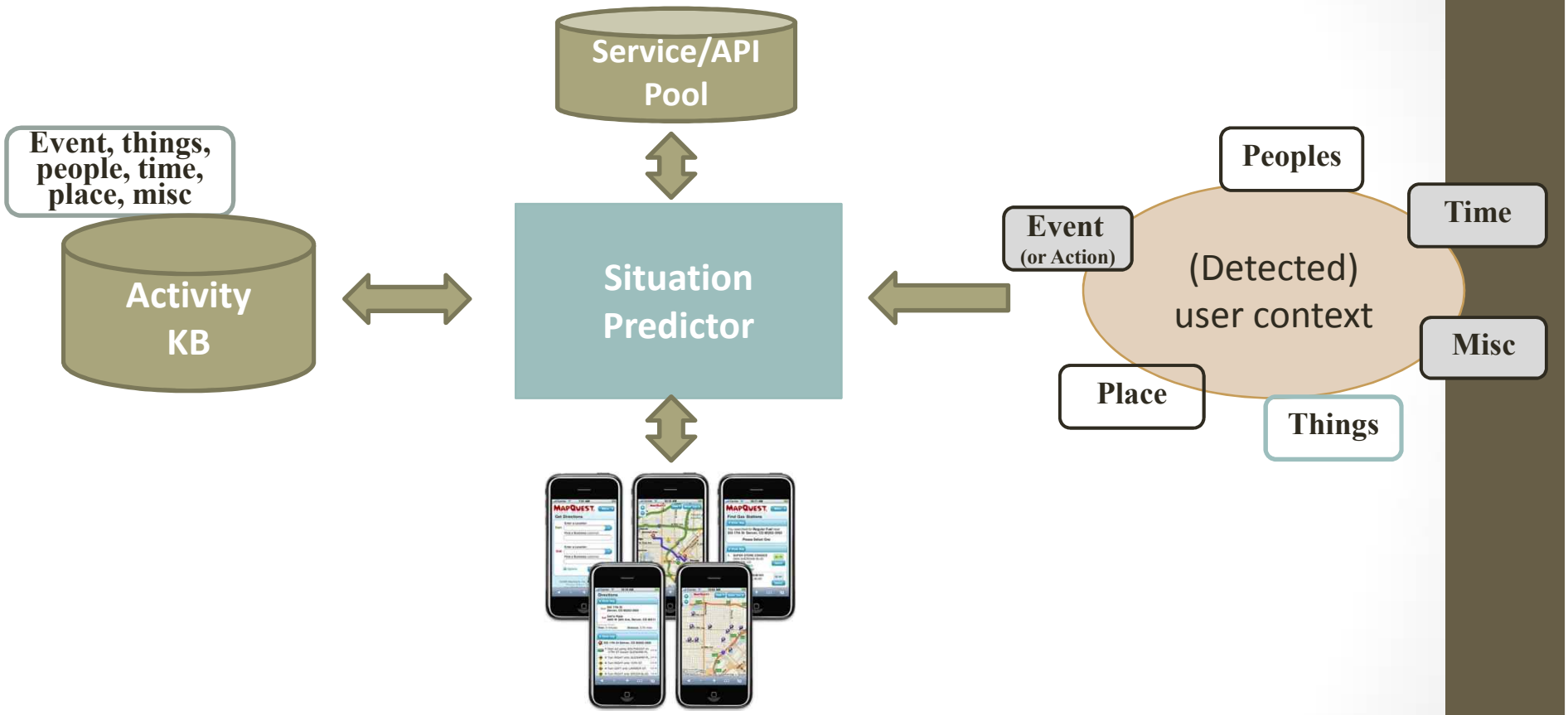
Query Free Search



- **Query free search without any manual query**

Excerpted from [Y. C. Jung, 2010]

Situation Prediction



- **Based on the Situation Predictor using AKB**
  - Recommend possible services: (android) app. applications
  - Recommend web service APIs for semantic web service composition

Excerpted from [Y. C. Jung, 2010]

Concluding Remarks

- For more complete event extractions for unstructured text
  - Needs to effectively utilize multiple knowledge resources (e.g., Freebase, Wikipedia, etc.)
  - Temporal, spatial, topical analysis is required

- Event-based knowledge base construction is a way to build situational knowledge that is essential for context-aware applications.
  - Semantic search, service recommendation, intent analysis are promising candidates

References (1/2)

- Alan Ritter, Mausam, Oren Etzioni, Sam Clark, “Open Domain Event Extraction from Twitter”, Proc. of KDD 2012, pp. 1104-1112.
- Lexing Xie, Hari Sundaram, and Murray Cambell, “Event Mining in Multimedia Streams”, Proc. of the IEEE, vol. 96, no. 4, April 2008, pp. 623-647.
- Andrew S. Gordon, “Mining Commonsense Knowledge from Personal Stories in Internet Weblogs”, Proc. of the First Workshop on Automated Knowledge Base Construction, 2010.
- Andrew S. Gordon, “The Design of Knowledge-rich Browsing Interfaces for Retrieval in Digital Libraries”, Northwestern University Ph.D. Dissertation, Department of Computer Science, 1999.
- Jose Espinosa and Henry Lieberman, “EventNet: Inferring Temporal Relations Between Commonsense Events”, Proc. of MICAI 2005, pp. 61-69.
- Ting Liu and Tomek Strzalkowski, “Bootstrapping Events and Relations from Text”, Proc. of EACL 2012, pp. 296-305.
- Roser Sauri, Rober Knippen, et al., “Evita: A Robust Event Recognizer for QA Systems”, Proc. of HLT-EMNLP 2005, pp. 700-707.
- Libin Hou, Peifeing Li, et al., “Event Argument Extraction based on CRF”, Proc. of CLSW 2012, LNAI7717, pp. 32-39.
- Jethro Borsje, Frederik Hogenboom, and Flavius Frasincar, “Semi-automatic Financial Events Discovery based on Lexico-semantic Patterns”, Int. J. Web Engineering and Technology, vol. 6, no. 2, pp. 115-140.

References (2/2)

- Souleiman Hasan, Sean O'Riain, and Edward Curry, "Approximate Semantic Matching of Heterogeneous Events", Proc. of DEBS'12, pp. 252-263.
- Giang Binh Tran, Mohammad Alrifai, and Dat Quoc Nguyen, "Predicting Relevant News Events for Timeline Summaries", Proc. of WWW (Companion Volume) 2013, pp. 91-92.
- Sheng-Hao Hung, Chia-Hung Lin, and Jen-Shin Hong, "Web mining for event-based commonsense knowledge using lexico-syntactic pattern matching and semantic role labeling", Expert System with Applications 37, 2010, pp. 341-347.
- Yuchul Jung, "An Activity Knowledge Base from Multiple Resources: Construction and Utilizations", Korea Advanced Institute Science and Technology, Ph.D. Dissertation, Department of Computer Science, 2010.
- Jannik Strotgen and Michael Gertz, "Event-centric Search and Exploration in Document Collections", Proc. of JCDL'12, pp. 223-232.
- Tomas Lin et al., "Active Objects: Actions for Entity-Centric Search", Proc. of WWW 2012, pp. 589-598.