Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts

Jiwei Li[1], Alan Ritter[2], Claire Cardie[3] and Eduard Hovy[4]

Department of Computer Science
[1]Stanford University [2]Ohio State University
[3]Cornell University [4]Carnegie Mellon University

June 22nd, 2014

Life Events

Life Events on Social Media

Life Events on Social Media

Jessica Jones @jonesalgebra · Sep 27

We're **engaged**!!!! I could not be more thrilled! We are getting married June 12, 2015!

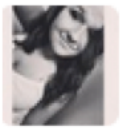
View photo

Reply Retweet Favorite More

Life Events on Social Media

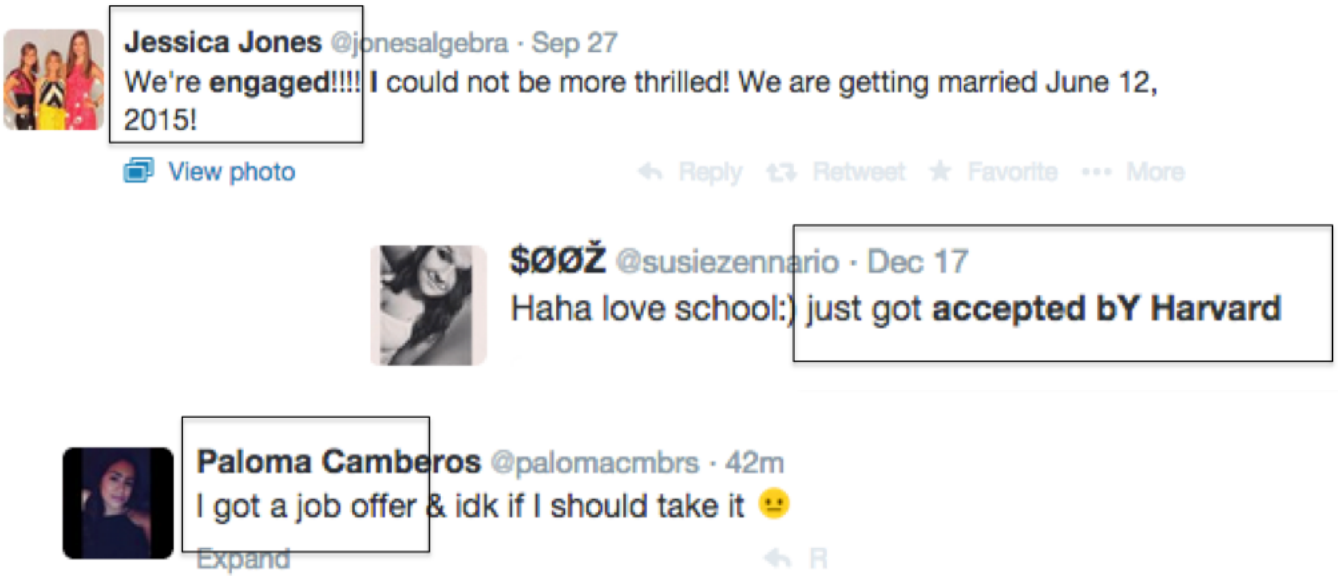
Jessica Jones @jonesalgebra · Sep 27
We're **engaged**!!!! I could not be more thrilled! We are getting married June 12, 2015!

View photo Reply Retweet Favorite More



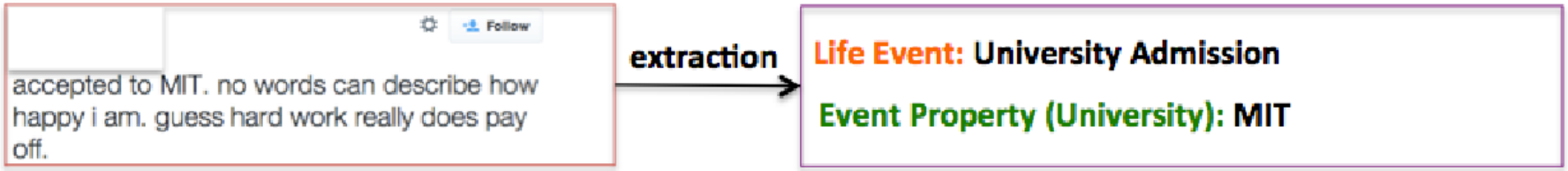
$ØØŽ @susiezennario · Dec 17
Haha love school:) just got **accepted bY Harvard**

Life Events on Social Media

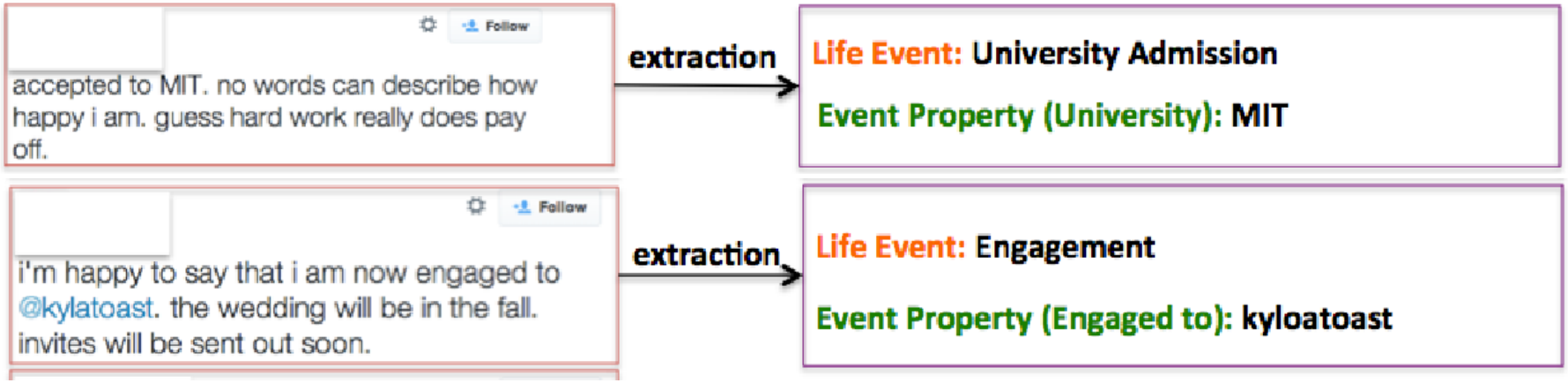


Life Events on Social Media

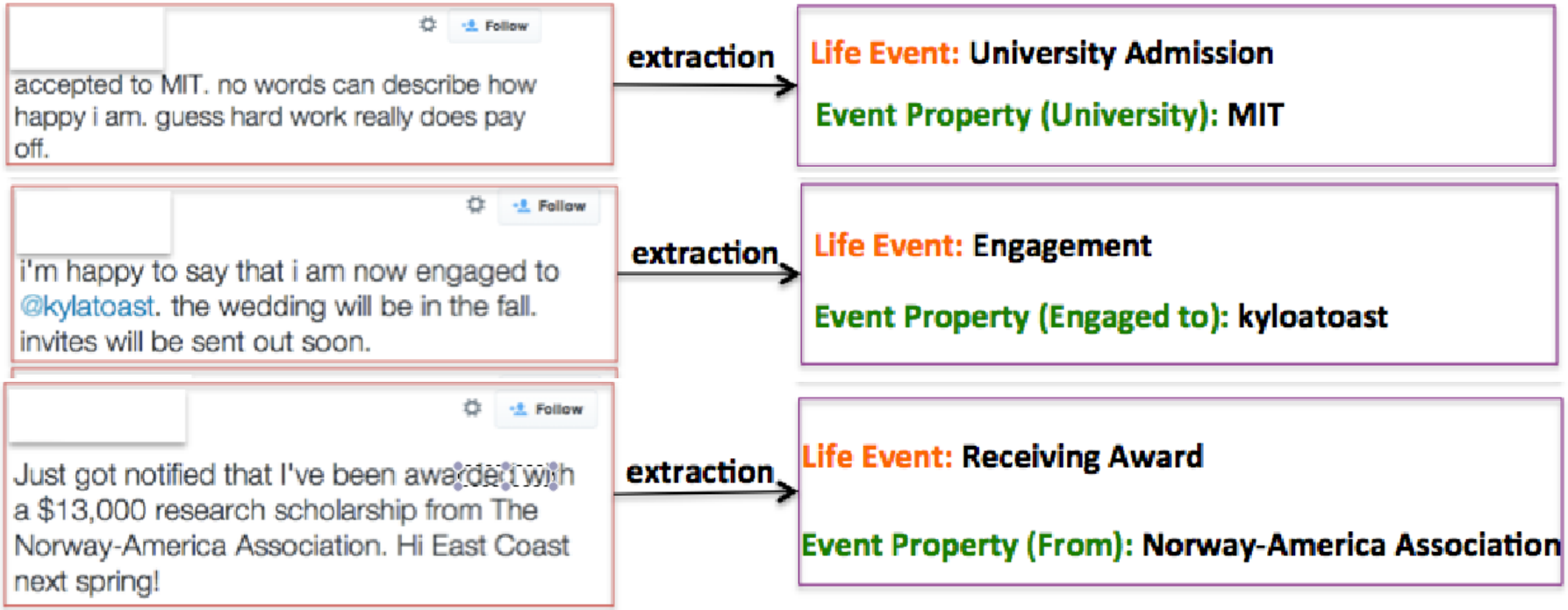
Life Events on Social Media



Life Events on Social Media



Life Events on Social Media



Life Events on Social Media

Why ?

Life Events on Social Media

Why ?

- Better understanding of users

Life Events on Social Media

Why ?

- Better understanding of users
- Friend Recommendation

Life Events on Social Media

Why ?

- Better understanding of users
- Friend Recommendation
- Online advertising

Outline

- Challenges
- System Overview
- Algorithms
- Experiments
- Conclusion

- Challenges

Challenges

Challenge 1: Major life event is an ambiguous concept !

Challenges

Challenge 1: Major life event is an ambiguous concept !

Challenges

Challenge 1: Major life event is an ambiguous concept !

Challenges

Challenge 1: Major life event is an ambiguous concept !

Challenges

Challenge 1: Major life event is an ambiguous concept !

Challenges

Challenge 1: What are life events ?

Challenges

Challenge 1: What are life events ?

Challenges

Challenge 1: What are life events ?

Challenges

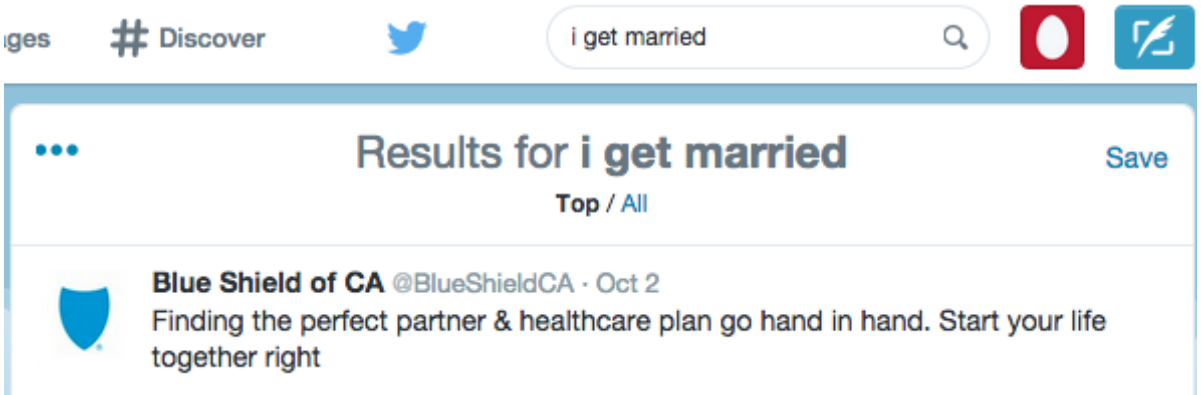
Challenge 1: What are life events ?

Challenges

Challenge 2: Noisy Data

Challenges

Challenge 2: Noisy Data



Challenges

Challenge 2: Noisy Data

Challenges

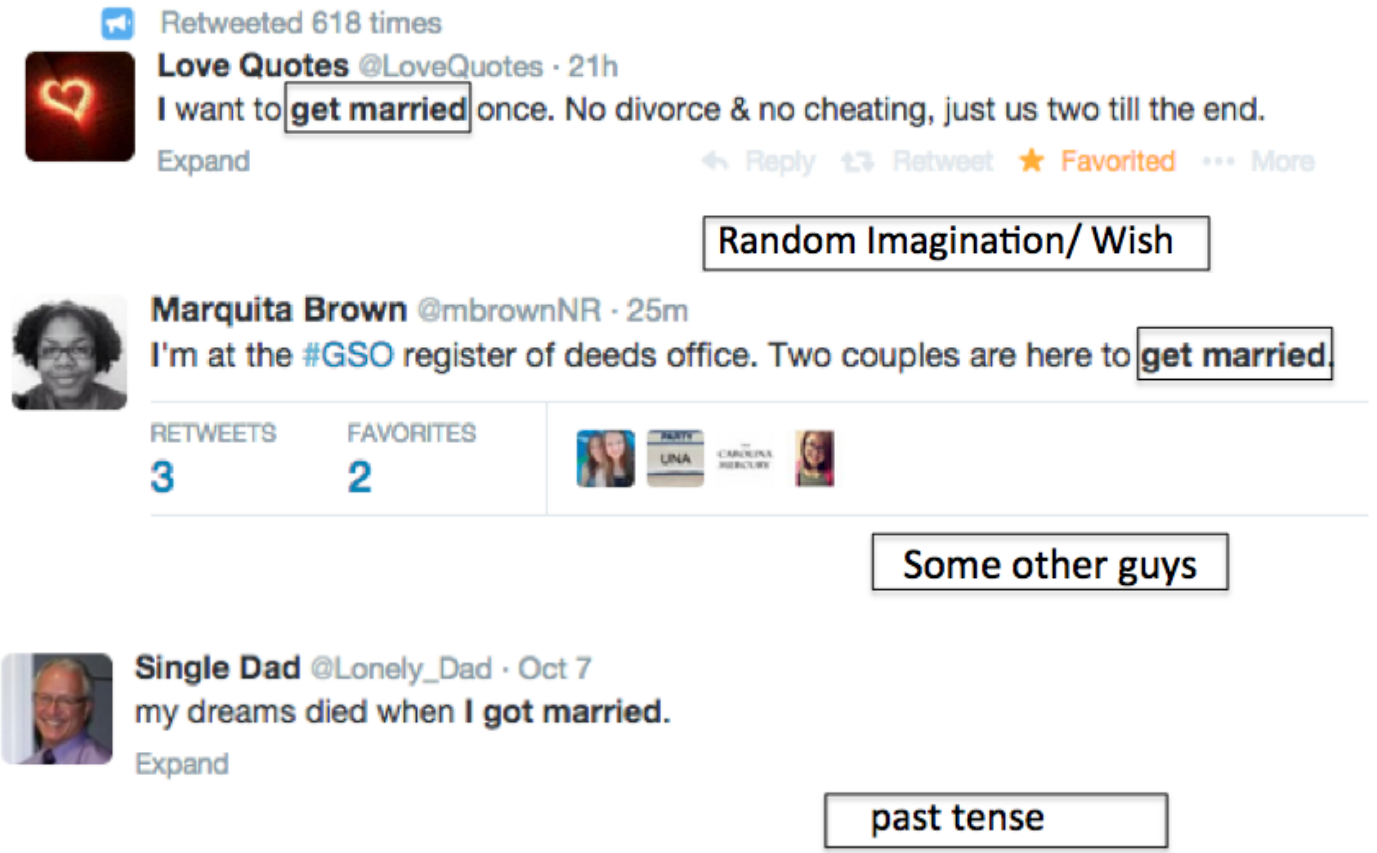
Challenge 2: Noisy Data

Challenges

Challenge 2: Noisy Data

Challenges

Challenge 2: Noisy Data



Challenges

Challenge 3: Lack of labeled data

Challenges

Challenge 3: Lack of labeled data

- No labeling criteria

Challenges

Challenge 3: Lack of labeled data

- No labeling criteria
- Life events sparsely distributed

Challenges

Challenge 3: Lack of labeled data

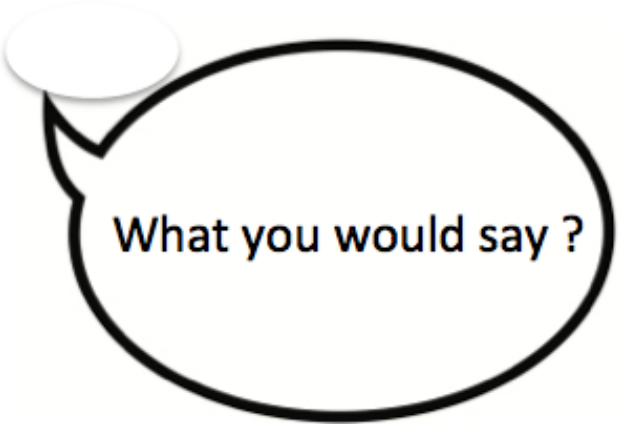
- No labeling criteria
- Life events sparsely distributed
- Rare events

Challenges

HOW ? ?

Challenges

Challenges



Challenges

Challenges

Challenges


Congratulations!


great!


Fantastic!


Awesome


COOL!


I'm so
sorry to
hear that.


"THAT'S
TERRIBLE"

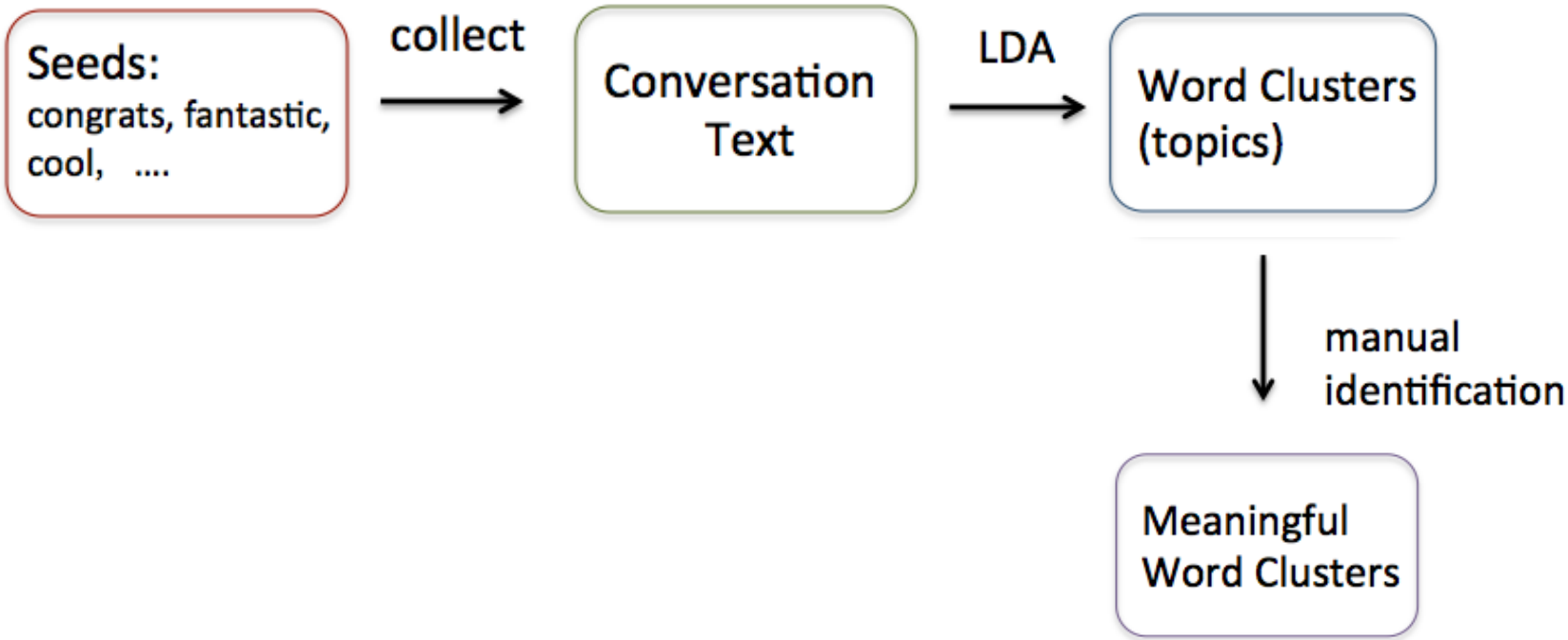
Responses based Data Harvesting

Seeds:
congrats, fantastic,
cool, ...

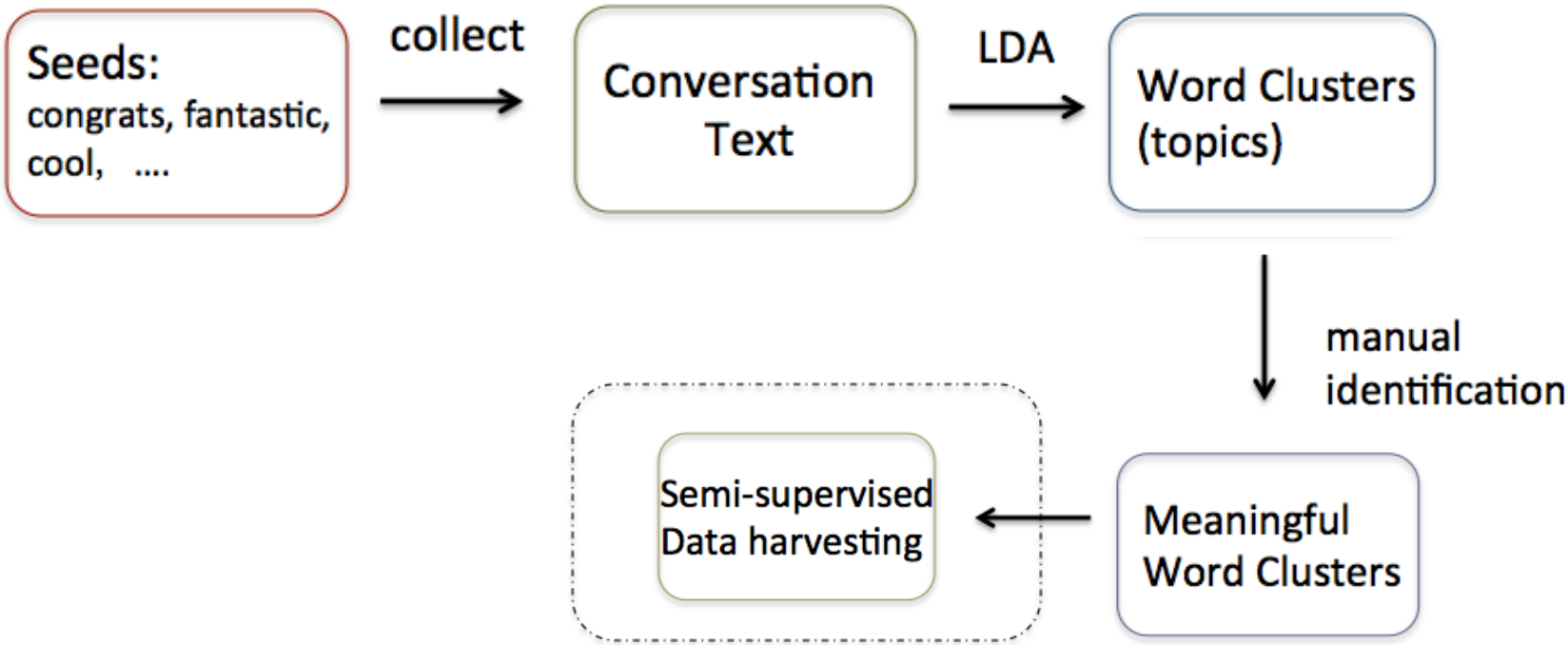
Responses based Data Harvesting

Responses based Data Harvesting

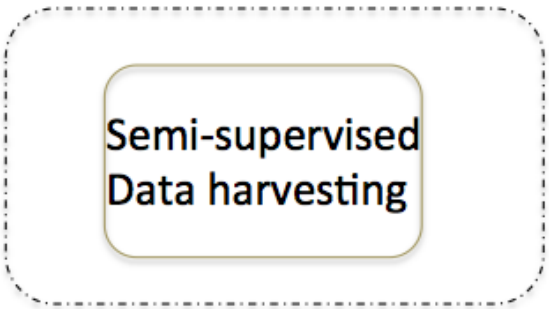
Responses based Data Harvesting



Responses based Data Harvesting



Responses based Data Harvesting

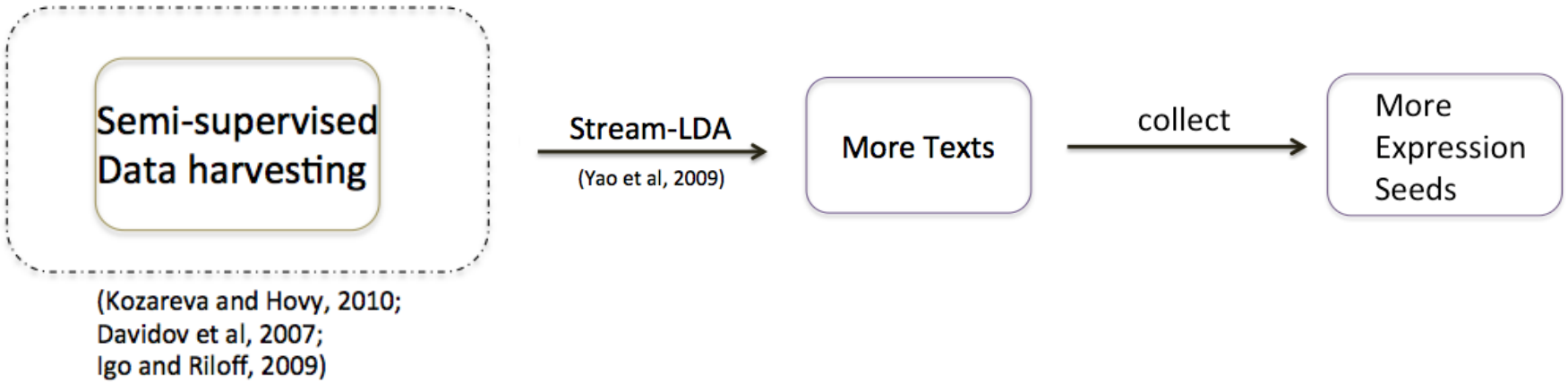


(Kozareva and Hovy, 2010;
Davidov et al, 2007;
Igo and Riloff, 2009)

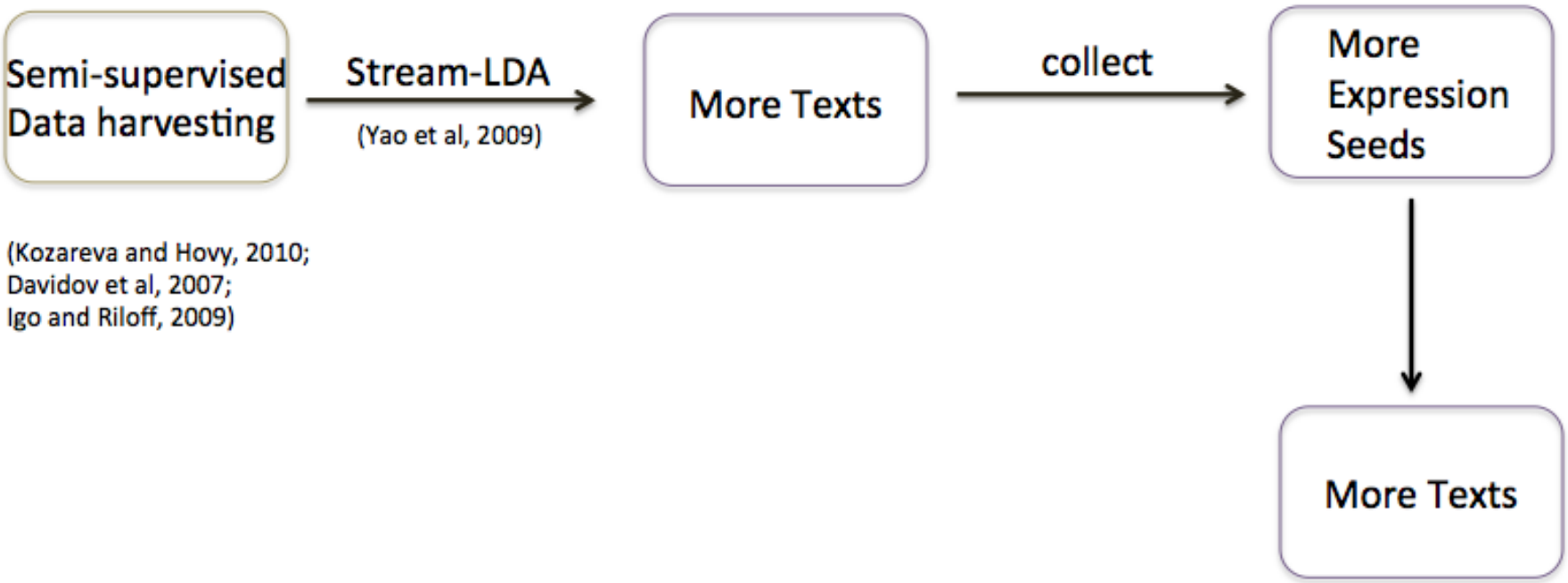
Responses based Data Harvesting

(Kozareva and Hovy, 2010;
Davidov et al, 2007;
Igo and Riloff, 2009)

Responses based Data Harvesting

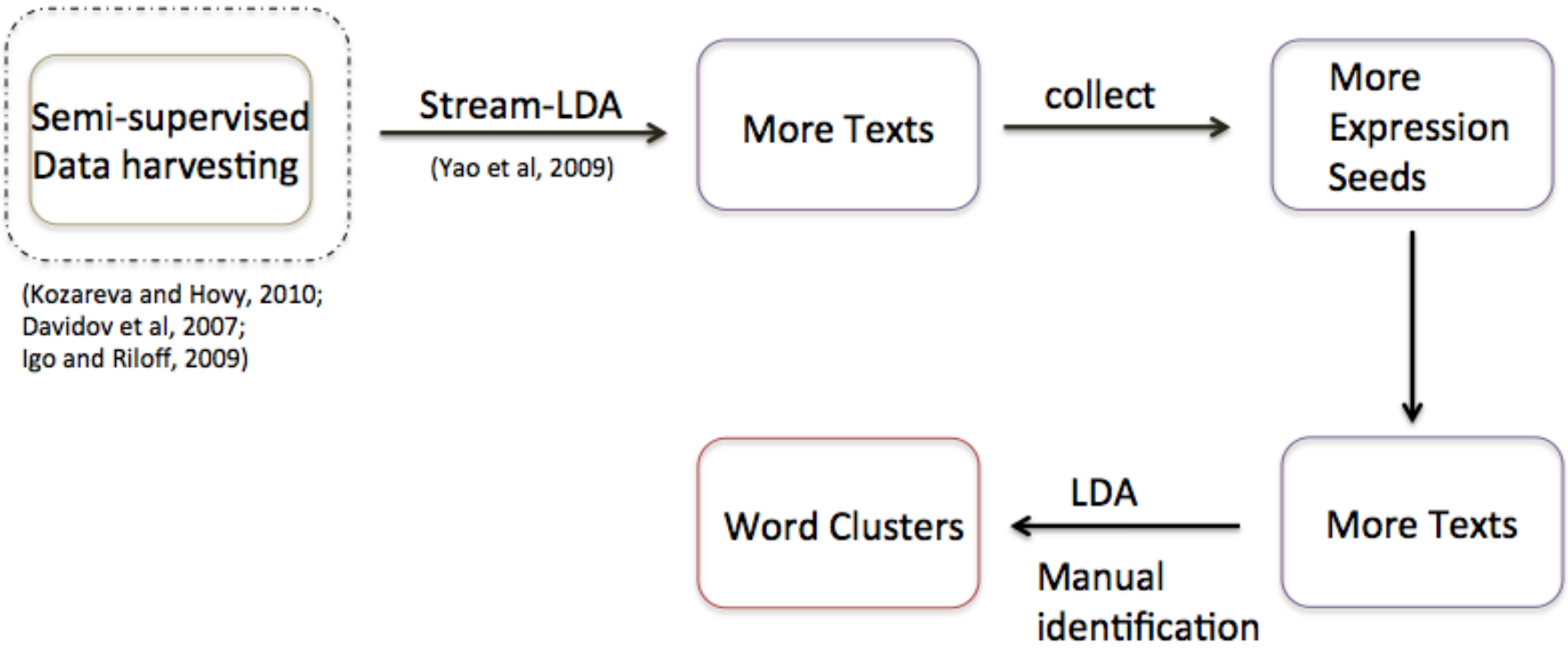


Responses based Data Harvesting

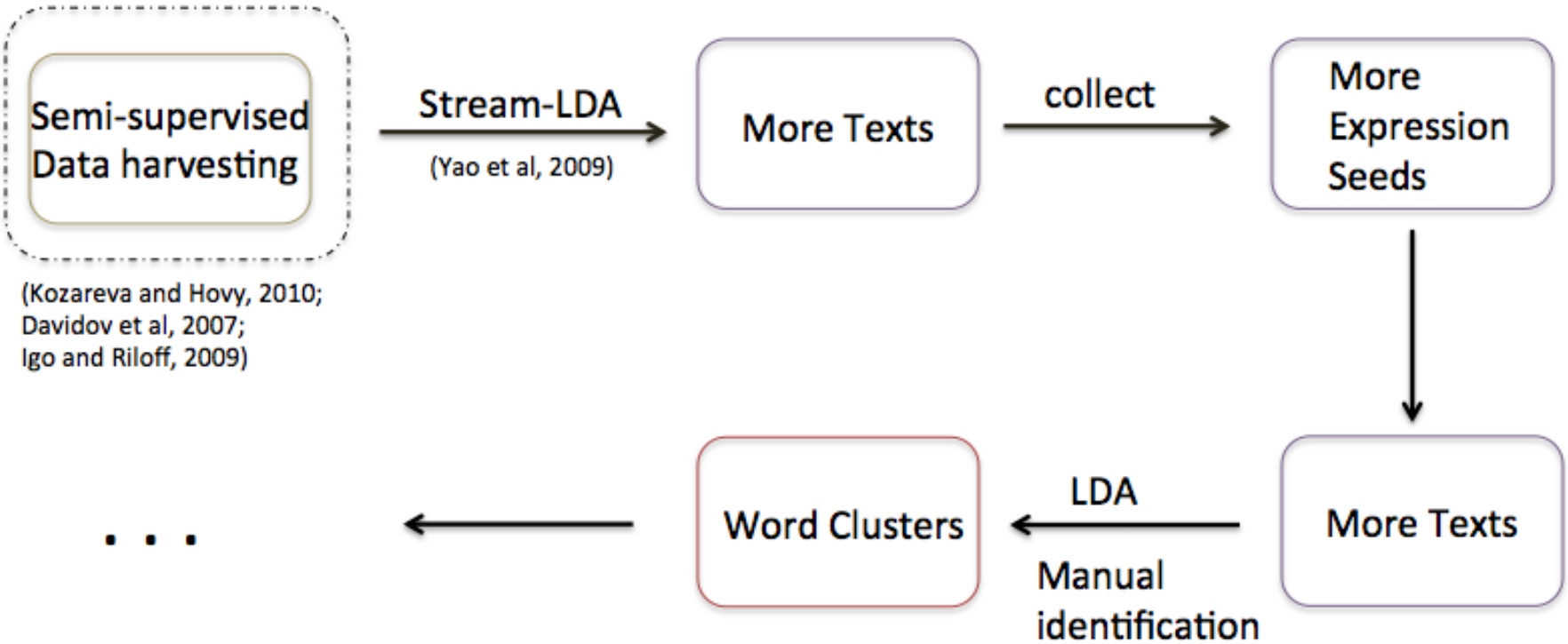


(Kozareva and Hovy, 2010;
Davidov et al, 2007;
Igo and Riloff, 2009)

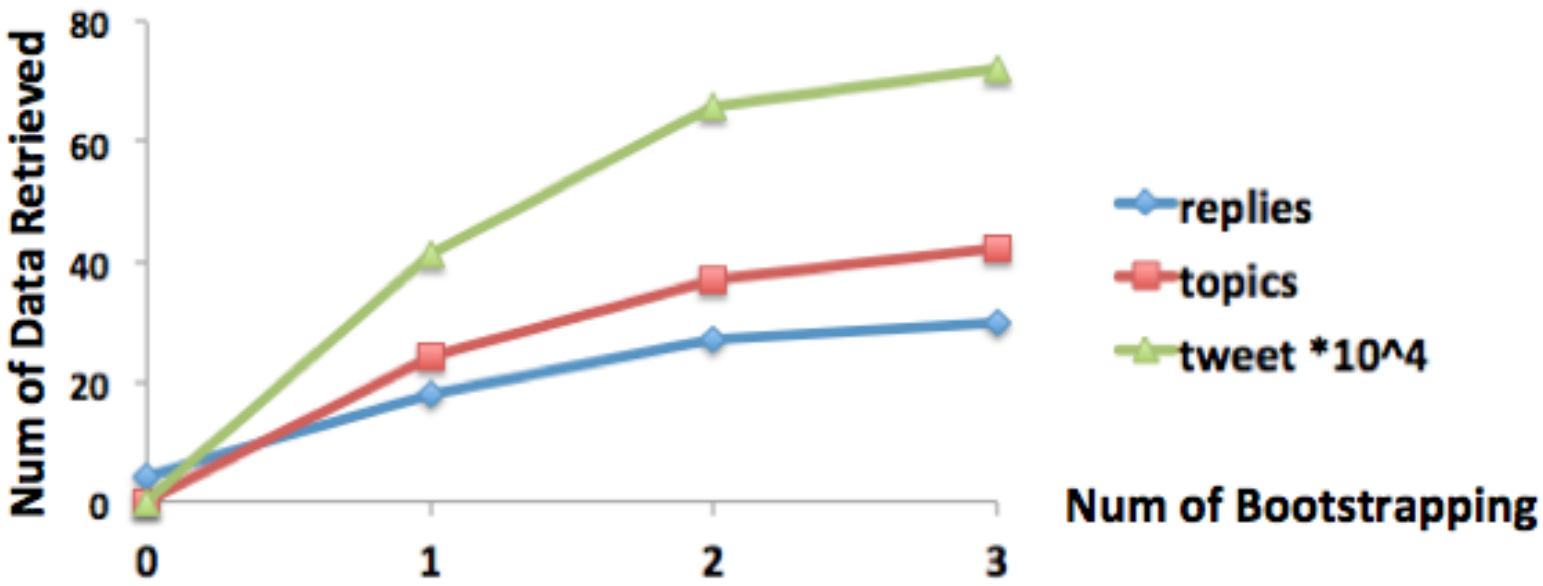
Responses based Data Harvesting



Responses based Data Harvesting



Responses based Data Harvesting



Responses based Data Harvesting

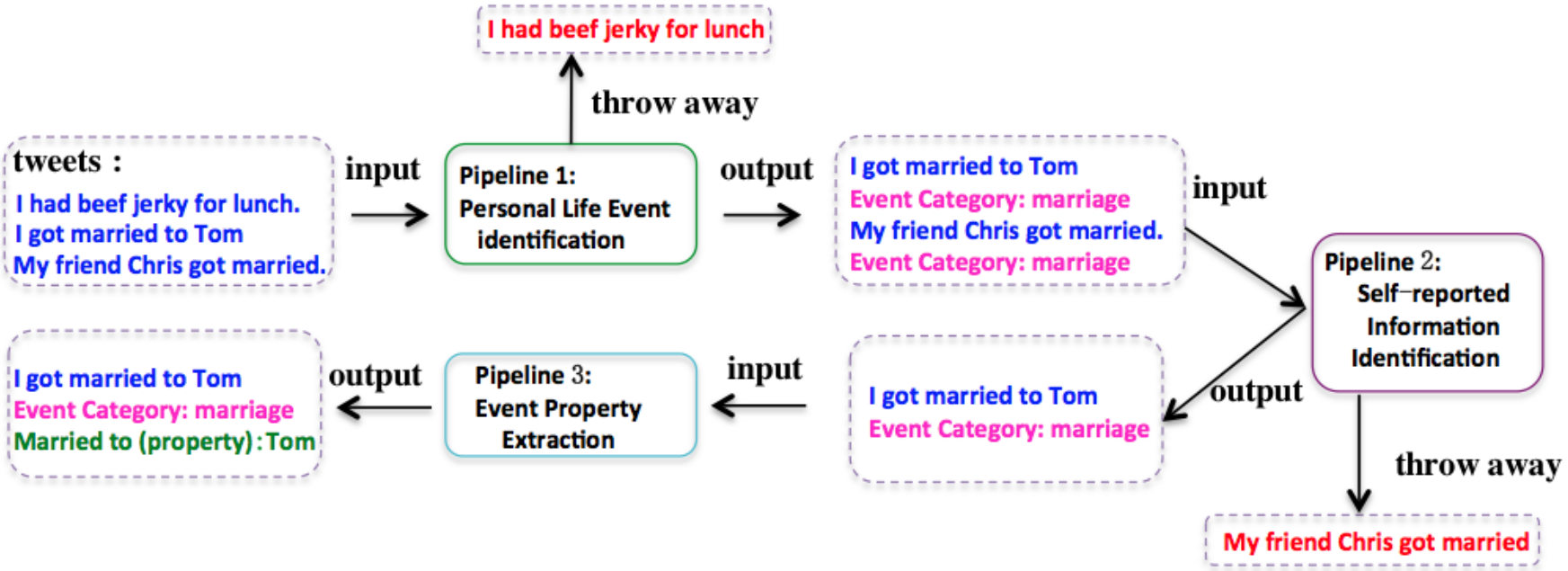
Life Event	Proportion	Life Event	Proportion
Birthday	9.78	Vacation	2.24
Job	8.39	Relationship	2.16
Wedding	7.24	Exams	2.02
Award	6.20	Election	1.85
Sports	6.08	New Car	1.65
Anniversary	5.44	Running	1.42
Give Birth	4.28	Surgery	1.20
Graduate	3.86	Lawsuit	0.64
Death	3.80	Acting	0.50
Admission	3.54	Research	0.48
Interview	3.44	Essay	0.35
Moving	3.26	Lost Weight	0.35
Travel	3.24	Publishing	0.28
Illness	2.45	Song	0.22

Table 1 : List of automatically discovered life event types.

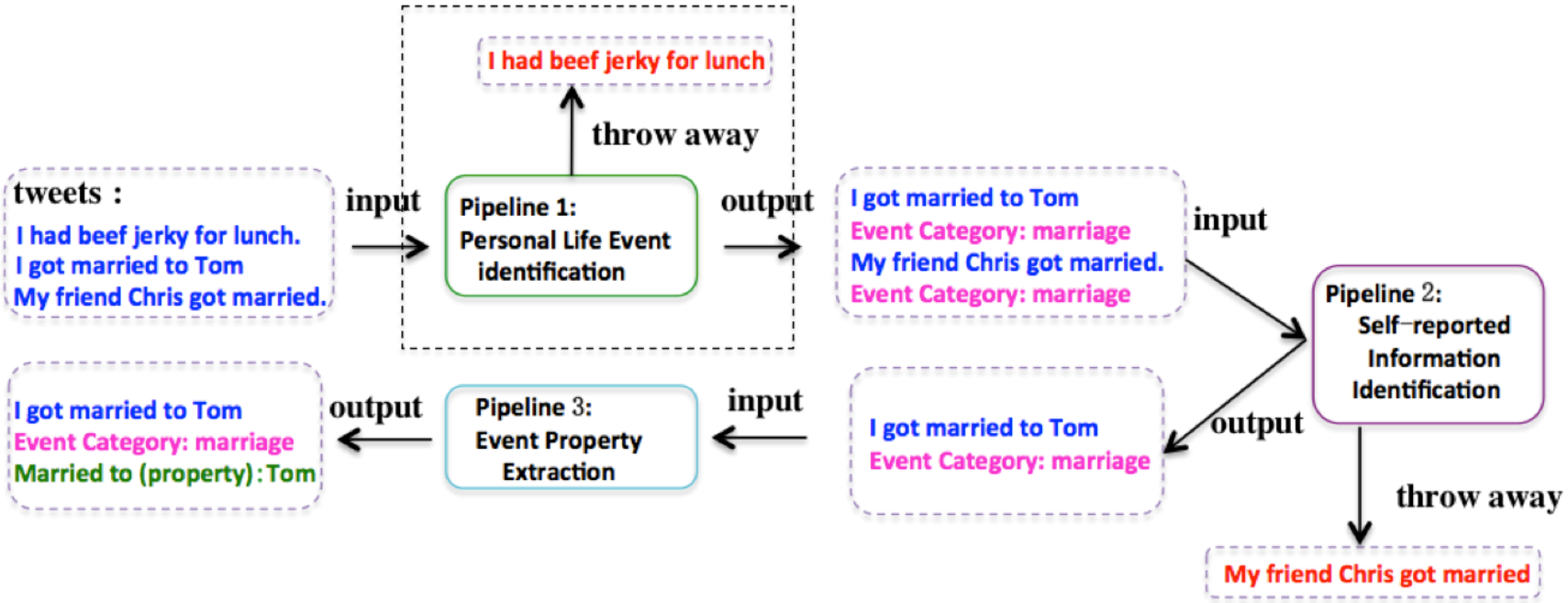
Responses based Data Harvesting

Human Label	Top words
Wedding	wedding, love, ring, engagement, engaged, bride, video, marrying
Graduation	graduation, school, college, graduate, graduating, year, grad
Relationship	boyfriend, girlfriend, date, check, relationship, see, look
Anniversary	anniversary, years, year, married, celebrating, wife, celebrate, love
Admission	admitted, university, admission, accepted, college, offer, school
Exam	passed, exam, test, school, semester, finished, exams, midterms
Research	research, presentation, journalism, paper, conference, go, writing
Essay & Thesis	essay, thesis, reading, statement, dissertation, complete, project
Job	job, accepted, announce, join, joining, offer, starting, announced, work
Interview	interview, position, accepted, internship, offered, start, work
Moving	house, moving, move, city, home, car, place, apartment, town, leaving
Travel	leave, leaving, flight, home, miss, house, airport, packing, morning
Vacation	vocation, family, trip, country, go, flying, visited, holiday, Hawaii
Winning Award	won, award, support, awards, winning, honor, scholarship, prize
Election	president, elected, run, nominated, named, promotion, cel, selected, business, vote
Publishing	book, sold, writing, finished, read, copy, review, release, books, cover
Contract	signed, contract, deal, agreements, agreed, produce, dollar, meeting
song	video, song, album, check, show, see, making, radio, love
Acting	play, role, acting, drama, played, series, movie, actor, theater
Death	dies, passed, cancer, family, hospital, dad, grandma, mom, grandpa
Give Birth	baby, born, boy, pregnant, girl, lbs, name, son, world, daughter, birth
Illness	ill, hospital, feeling, sick, cold, flu, getting, fever, doctors, cough
Surgery	surgery, got, test, emergency, blood, tumor, stomachs, hospital, pain, brain
Sports	win, game, team, season, fans, played, winning, football, luck
Running	run, race, finished, race, marathon, ran, miles, running, finish, goal
New Car	car, buy, bought, cars, get, drive, pick, seat, color, dollar, meet
Lost Weight	weight, lost, week, pounds, loss, weeks, gym, exercise, running

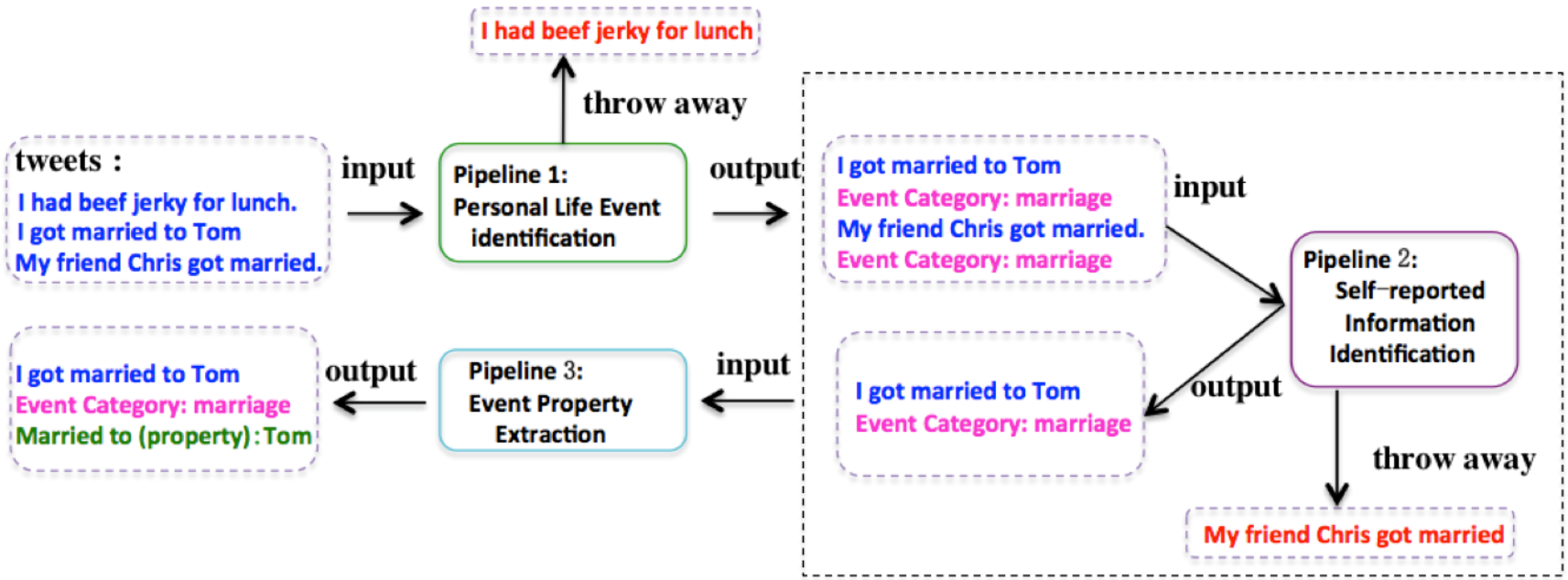
System Overview



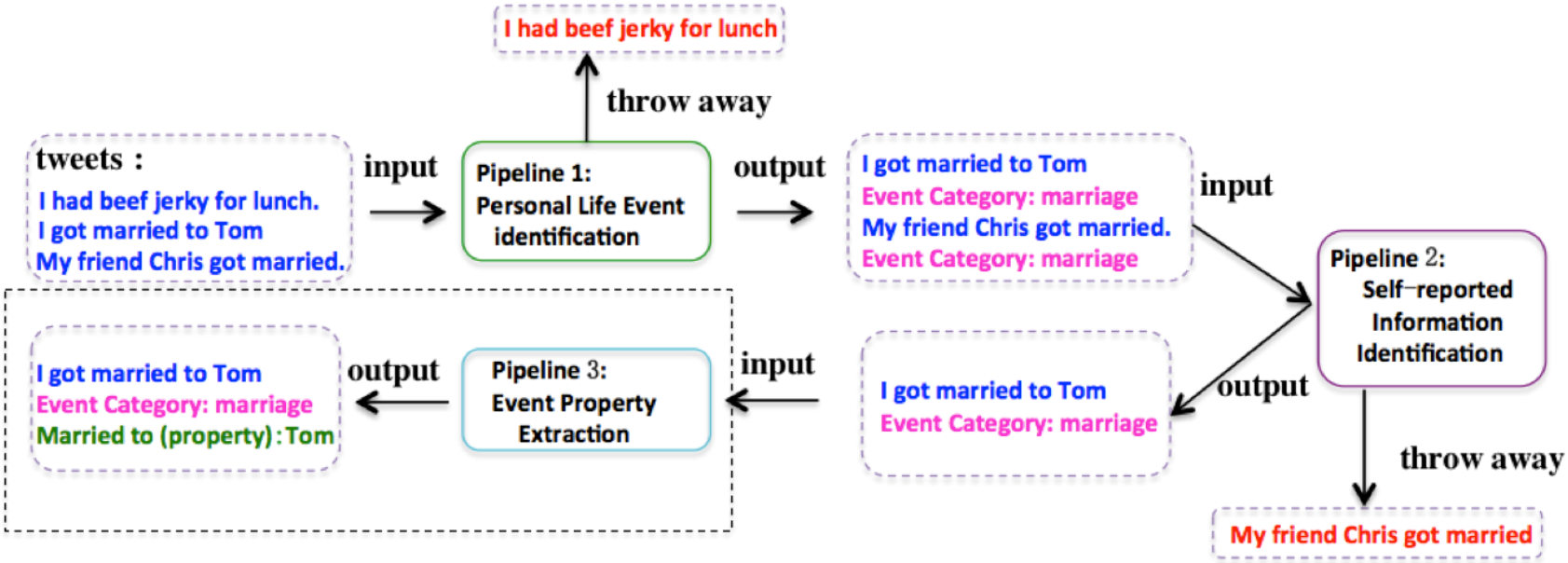
System Overview



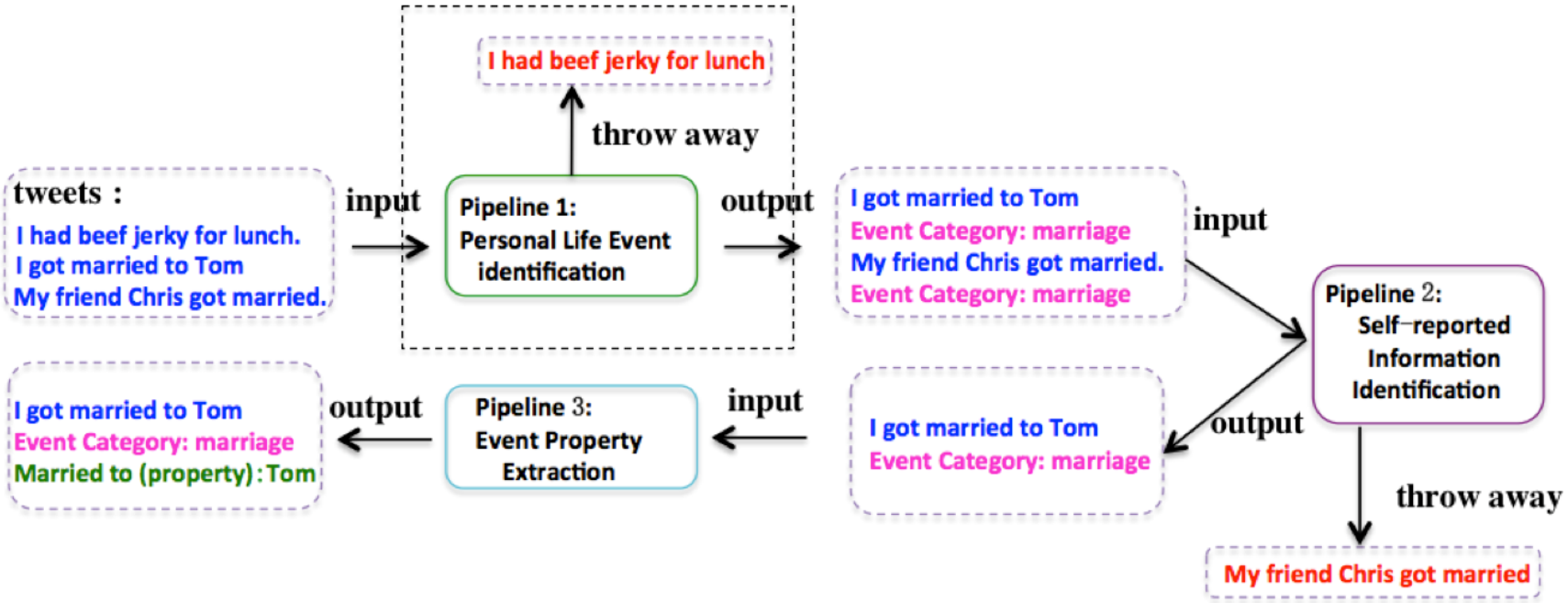
System Overview



System Overview



Personal Event Identification



Personal Event Identification

Multi-Class Classifier based on SVM

Personal Event Identification

Multi-Class Classifier based on SVM

Positive Examples for each category: Pre-identified data

Personal Event Identification

Multi-Class Classifier based on SVM

Positive Examples for each category: Pre-identified data

Negative Examples: Random Tweets

Personal Event Identification

Multi-Class Classifier based on SVM

Positive Examples for each category: Pre-identified data

Negative Examples: Random Tweets

Personal Event Identification

Multi-Class Classifier based on SVM

Positive Examples for each category: Pre-identified data

Negative Examples: Random Tweets

- Topic-Tweet probability

Personal Event Identification

Multi-Class Classifier based on SVM

Positive Examples for each category: Pre-identified data

Negative Examples: Random Tweets

- Topic-Tweet probability
- Dictionary

Personal Event Identification

Multi-Class Classifier based on SVM

Positive Examples for each category: Pre-identified data

Negative Examples: Random Tweets

- Topic-Tweet probability
- Dictionary
- Word, NER, POS
- Window Context

Personal Event Identification

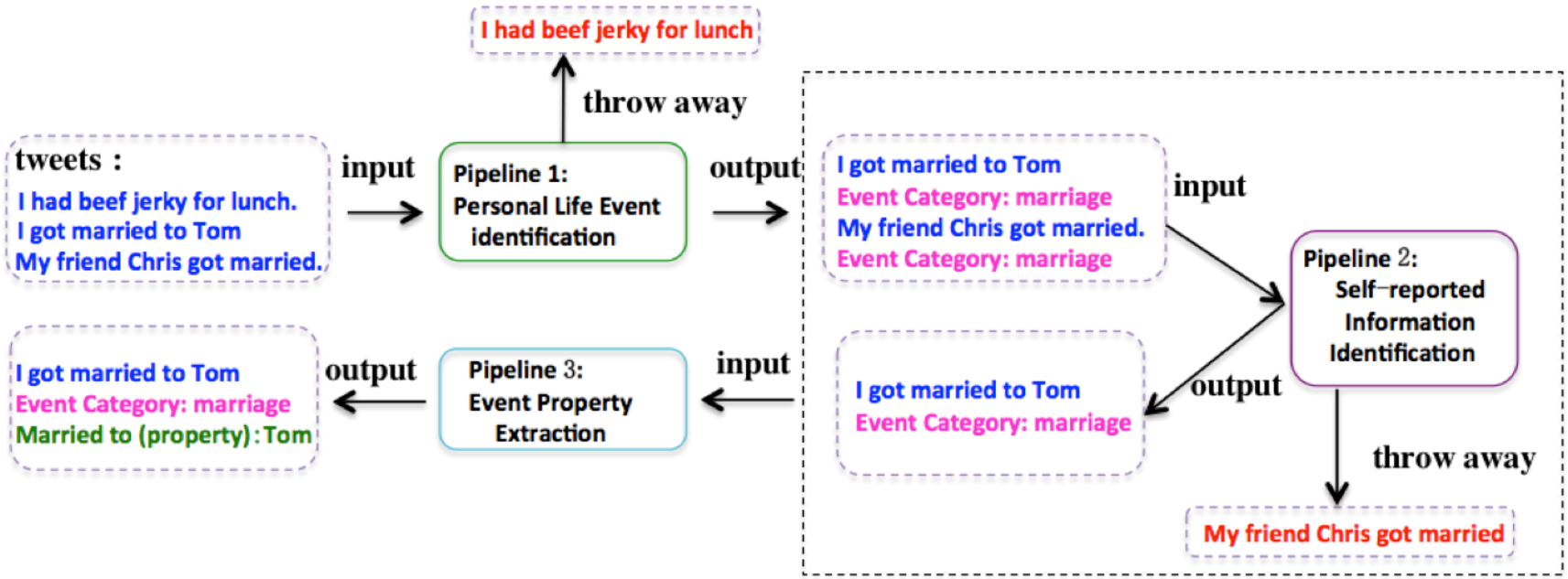
Multi-Class Classifier based on SVM:

Split harvested data, training and testing

Feature Setting	Precision	Recall
Word+NER	0.204	0.326
Word+NER+Dictionary	0.362	0.433
All	0.382	0.487

Self Information Identification

Self Information Identification



Self Information Identification

Negative Examples

Self Information Identification

Negative Examples

- Not self

Self Information Identification

Negative Examples

- Not self
- Random Thought

Self Information Identification

Negative Examples

- Not self
- Random Thought
- Past Tense

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Binary SVM Classifier

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Binary SVM Classifier

- Tense

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Binary SVM Classifier

- Tense
- Factuality (could, would, can ...) (Saurf and Pustejovsky, 2007)

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Binary SVM Classifier

- Tense
- Factuality (could, would, can ...) (Saurf and Pustejovsky, 2007)
- I

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Binary SVM Classifier

- Tense
- Factuality (could, would, can ...) (Saurf and Pustejovsky, 2007)
- I
- Dependency (Kong et al., 2014)

Self Information Identification

Dataset:
Positive: selected from harvested data
Negative: selected from harvested data

Binary SVM Classifier

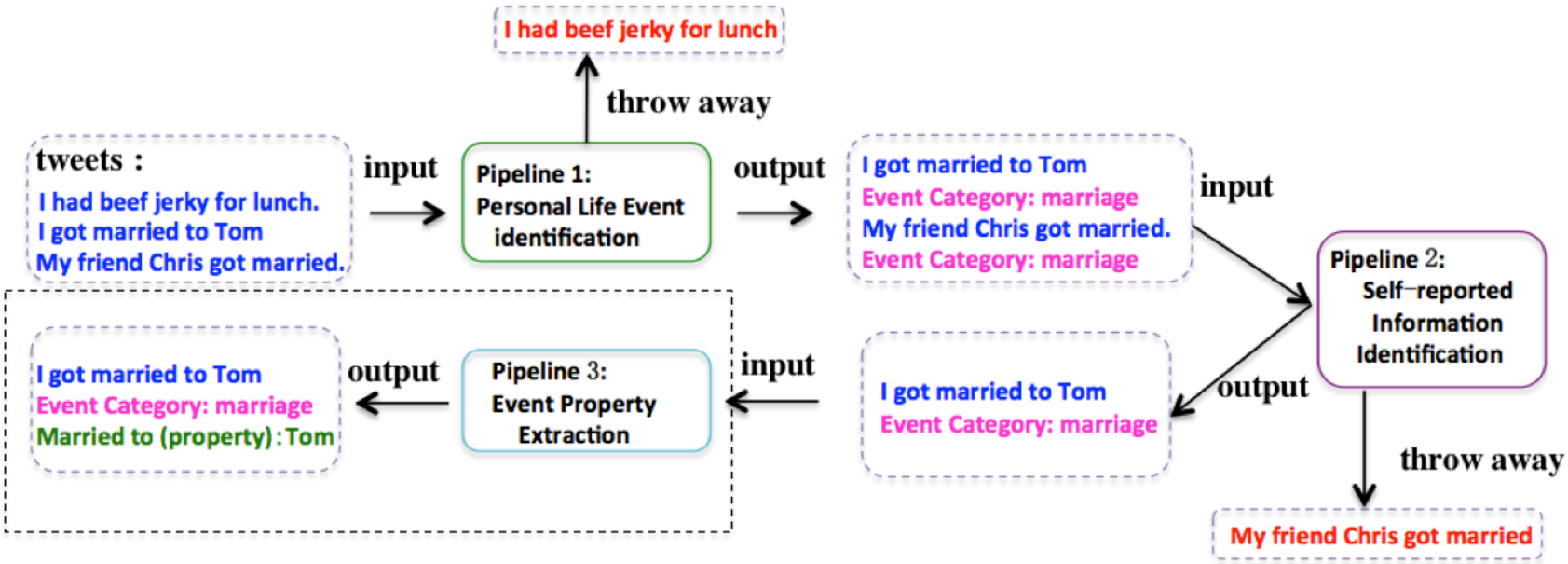
- Tense
- Factuality (could, would, can ...) (Saurf and Pustejovsky, 2007)
- I
- Dependency (Kong et al., 2014)
- Token, NER, POS, window context

Self Information Identification

Feature Setting	Acc	Pre	Rec
Bigram+Window	0.76	0.47	0.44
Bigram+Window +Tense+Factuality	0.77	0.47	0.46
all	0.82	0.51	0.48

Event Property Identification

Event Property Identification



Event Property Identification

Human Labeling

Life Event	Property
(a) Acceptance, Graduation	Name of University/College
(b) Wedding, Engagement, Falling love	Name of Spouse/ partner/ bf/ gf
(c) Getting a job, interview, internship	Name of Enterprise
(d) Moving to New Places, Trip, Vocation, Leaving	Place, Origin, Destination
(e) Winning Award	Name of Award, Prize

Event Property Identification

Sequence Labeling Task, CRF (Lafferty, et al., 2001)

- Word token, Capitalization, POS, word shape, NER
- A gazetteer of universities and companies
- Context

System

What benefits brought from Congratulations/Condolences Speech Acts ?

- Clean Data

System

What benefits brought from Congratulations/Condolences Speech Acts ?

- Clean Data
  - Personal Topic Identification

System

What benefits brought from Congratulations/Condolences Speech Acts ?

- Clean Data
  - Personal Topic Identification
  - Self Report Information

System

What benefits brought from Congratulations/Condolences Speech Acts ?

- Clean Data
  - Personal Topic Identification
  - Self Report Information
    User 1: I wish to get married
    User 2: Congratulations ! !

Experiments

- End-to-End Experiments

Experiments

Gold-standard life event dataset

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets
- Ask Turkers to label other people's tweets.

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets
- Ask Turkers to label other people's tweets.
  - 2 Turkers 1 tweet

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets
- Ask Turkers to label other people's tweets.
  - 2 Turkers 1 tweet
  - Inter-rater agreement is 0.54 (cohen's kappa)

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets
- Ask Turkers to label other people's tweets.
  - 2 Turkers 1 tweet
  - Inter-rater agreement is 0.54 (cohen's kappa)
  - Authors make final decision

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets
- Ask Turkers to label other people's tweets.
  - 2 Turkers 1 tweet
  - Inter-rater agreement is 0.54 (cohen's kappa)
  - Authors make final decision
- 900 positive tweets

Experiments

Gold-standard life event dataset

- Ask Twitter users to label their own tweets
- Ask Turkers to label other people's tweets.
  - 2 Turkers 1 tweet
  - Inter-rater agreement is 0.54 (cohen's kappa)
  - Authors make final decision
- 900 positive tweets
- 60,000 negative tweets

Experiments

Baselines

Experiments

Baselines

- Supervised

Experiments

Baselines

- Supervised
- Supervised + Self

Experiments

Baselines

- Supervised
- Supervised + Self

Approach	Precision	Recall
Our approach	0.62	0.48
Supervised	0.13	0.20
Supervised+Self	0.25	0.18

Table 3 : Performance for different approaches for identifying life events.

Experiments

Does bootstrapping help ?

Experiments

Does bootstrapping help ?

Approach	Precision	Recall
Step 1	0.65	0.36
Step 2	0.64	0.43
Step 3	0.62	0.48

Table 4 : Performance for different steps of bootstrapping for identifying.

Conclusion

- Conclusion

Conclusion

Conclusion

- We study the life event extraction problem on Twitter

Conclusion

- We study the life event extraction problem on Twitter
- We propose a framework based on Congratulations/Condolences Speech Acts for data harvesting

Conclusion

- We study the life event extraction problem on Twitter
- We propose a framework based on Congratulations/Condolences Speech Acts for data harvesting
- We explore different types features and algorithms for this task

Conclusion

Key idea: solve this problem based on minimum human efforts.

Conclusion

Key idea: solve this problem based on minimum human efforts.

Problems

Conclusion

Key idea: solve this problem based on minimum human efforts.

Problems

- Restricted to event types identified by Congratulations/Condolences Speech Acts.

Conclusion

Key idea: solve this problem based on minimum human efforts.

Problems

- Restricted to event types identified by Congratulations/Condolences Speech Acts.
- No all responses correspond to life events

Conclusion

Key idea: solve this problem based on minimum human efforts.

Problems

- Restricted to event types identified by Congratulations/Condolences Speech Acts.
- No all responses correspond to life events
- Error accumulations.

Thank you !

Thank you !

Questions, Suggestions

Thank you !

Questions, Suggestions

Joint work with

Alan Ritter

Claire Cardie

Eduard Hovy