

# Towards semi-automated entity interaction annotation

Henry Feild  
(work with Phil Lombardo & Sam Alexander)

Endicott College  
December 6, 2018

# Outline

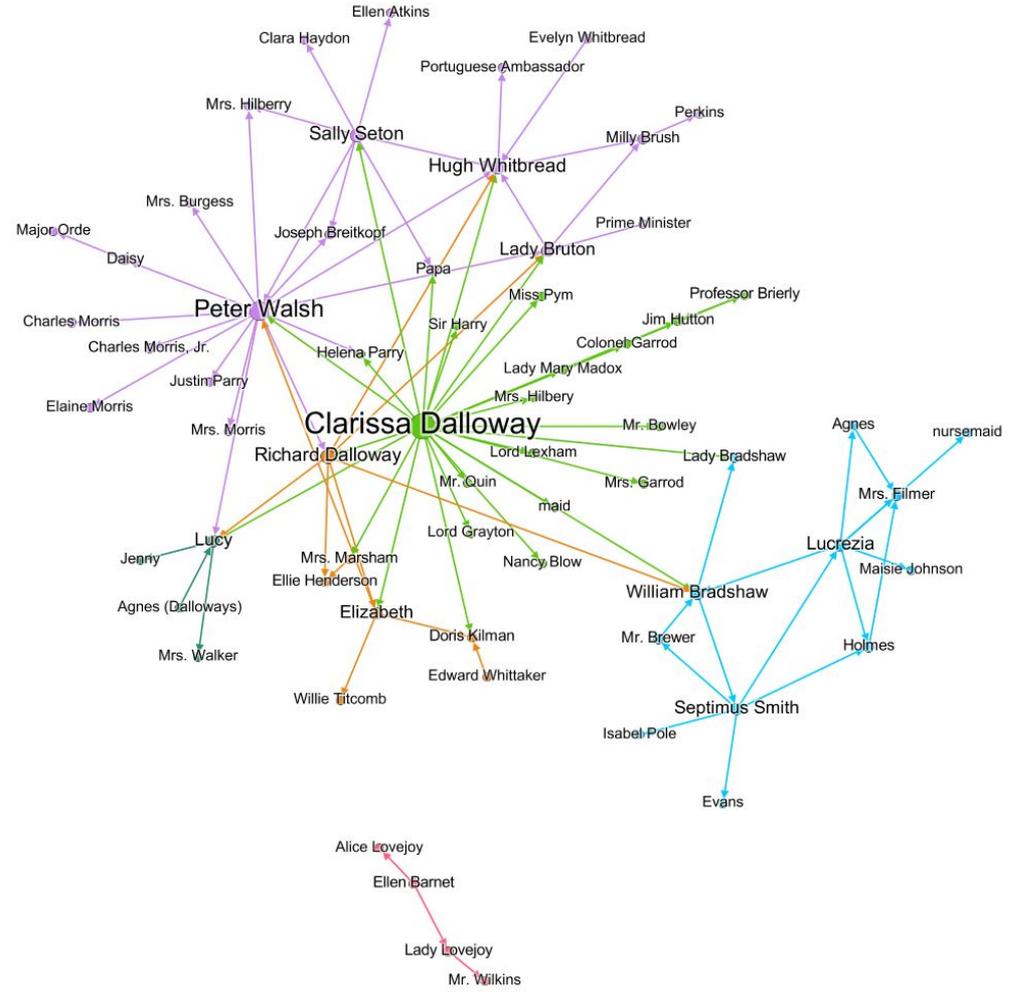
- about me
- motivation
- related work
- our vision
- what we have so far

# About me

- sixth year teaching at Endicott
  - in Beverly, MA (about 1 hour south of here)
    - in between witch trials and sad Casey Affleck movies
  - teaching school (4 courses per semester)
    - read: not much time for research :(
  - department of 2.5 faculty with ~40 undergrad CS majors
  - no graduate program
- before that: MS/PhD in Computer Science at UMass Amherst
  - Center for Intelligent Information Retrieval (along with Prof. Dietz, who was a postdoc)
  - overlapped with Prof. Petrik (we used to play volleyball and softball together)
- interned at NIST, Yahoo! (when it was still a thing), and MSR
  - definitely do an internship if you have a chance -- you meet amazing new people



# Motivation



# A page from Mrs. Dalloway by Virginia Woolf

"Heavens, the front-door bell," exclaimed Clarissa, staying her needle. Roused, she listened.

*pw-f*  
"Mrs. Dalloway will see me," said the elderly man in the hall. "Oh yes, she will see me," he repeated, putting Lucy aside very benevolently, and running upstairs ever so quickly. "Yes, yes, yes," he muttered as he ran upstairs. "She will see me. After five years in India, Clarissa will see me."

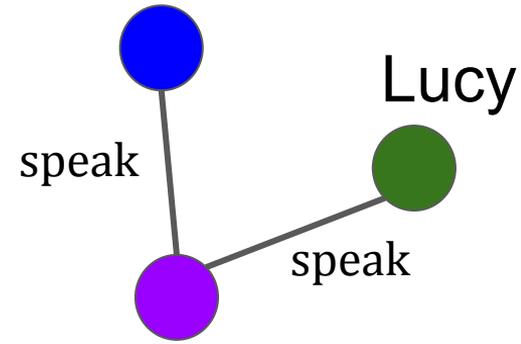
"Who can—what can," asked Mrs. Dalloway (thinking it was outrageous to be interrupted at eleven o'clock on the morning of the day she was giving a party), hearing a step on the stairs. She heard a hand upon the door. She made to hide her dress, like a virgin protecting chastity, respecting privacy. Now the brass knob slipped. Now the door opened, and in came—for a single second she could not remember what he was called! so surprised she was to see him, so glad, so shy, so utterly taken aback to have Peter Walsh come to her unexpectedly in the morning! (She had not read his letter.)

*pw-cd*  
"And how are you?" said Peter Walsh, positively trembling; taking both her hands; kissing both her hands. She's grown older, he thought, sitting down. I shan't tell her anything about it, he thought, for she's grown older. She's looking at me, he thought, a sudden embarrassment coming over him, though he had kissed her hands. Putting his hand into his pocket, he took out a large pocket-knife and half opened the blade.

Exactly the same, thought Clarissa; the same queer look; the same check suit; a little out of the straight his face is, a little thinner, dryer, perhaps, but he looks awfully well, and just the same.

"How heavenly it is to see you again!" she ex-

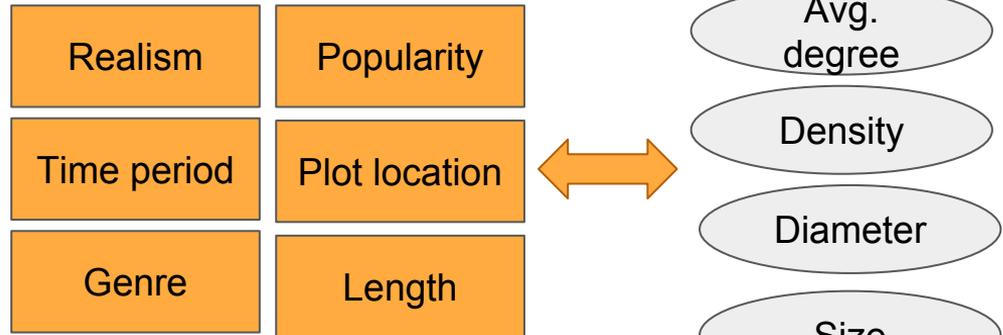
## Mrs. Dalloway



## Peter Walsh

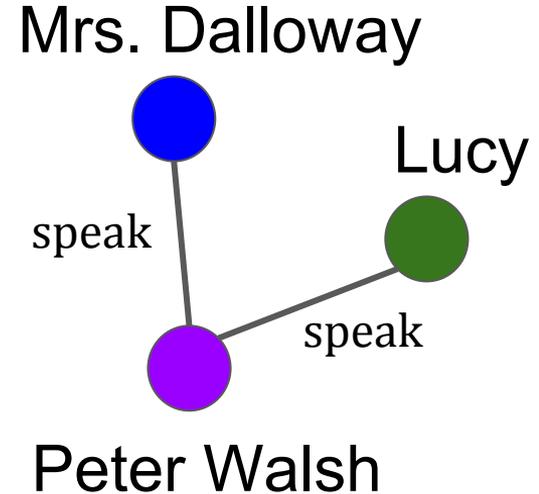
# Questions about social networks in literature

- how do conversation networks differ between rural and urban novels in 19th Century novels? (Elson..., 2010, 2012)
  - 60 novels analyzed using quoted speech attribution
  - (they don't)
- how do literary characteristics relate to character network characteristics? (Sack 2014; Alexander, unreleased)
- how big can a network get before readers can't keep track of characters anymore (Alexander)
- is there an inverse between amount of dialog and # of characters? (Elson et al., 2010; Agarwal 2016) (no, there's not)



# What's an edge?

- $x$  talks to  $y$
- $x$  observes  $y$
- $x$  knows about  $y$
- $x$  and  $y$  appear in the same scene together
- $x$  is related to  $y$
- others? probably



How can we get these annotations?

# Manual annotation pipeline

“Heavens, the ironing-press! Clarissa, staying her needle. Roused, she listened. “Mrs. Dalloway will see me,” said the elderly man in the hall. “Oh yes, she will see me,” he repeated, putting Lucy aside very benevolently, and running upstairs ever so quickly. “Yes, yes, yes,” he muttered as he ran upstairs. “She will see me. After five years in India, Clarissa will see me.”

“Who can—what can,” asked Mrs. Dalloway (thinking it was outrageous to be interrupted at eleven o’clock on the morning of the day she was giving a party), hearing a step on the stairs. She heard a hand upon the door. She made to hide her dress, like a virgin protecting chastity, respecting privacy. Now the brass knob slipped. Now the door opened, and in came—for a single second she could not remember what he was called! so surprised she was to see him, so glad, so shy, so utterly taken back to have Peter Walsh come to her unexpectedly in the morning! (She had not read his letter.)

“And how are you?” said Peter Walsh, positively trembling; taking both her hands; kissing both her hands. She’s grown older, he thought, sitting down. I shan’t tell her anything about it, he thought, for she’s grown older. She’s looking at me, he thought, a sudden embarrassment coming over him, though he had kissed her hands. Putting his hand into his pocket, he took out a large pocket-knife and half opened the blade.

Exactly the same, thought Clarissa; the same queer look; the same check suit; a little out of the straight his face is, a little thinner, dryer, perhaps, but he looks awfully well, and just the same.

“How heavenly it is to see you again!” she ex-

pw-f

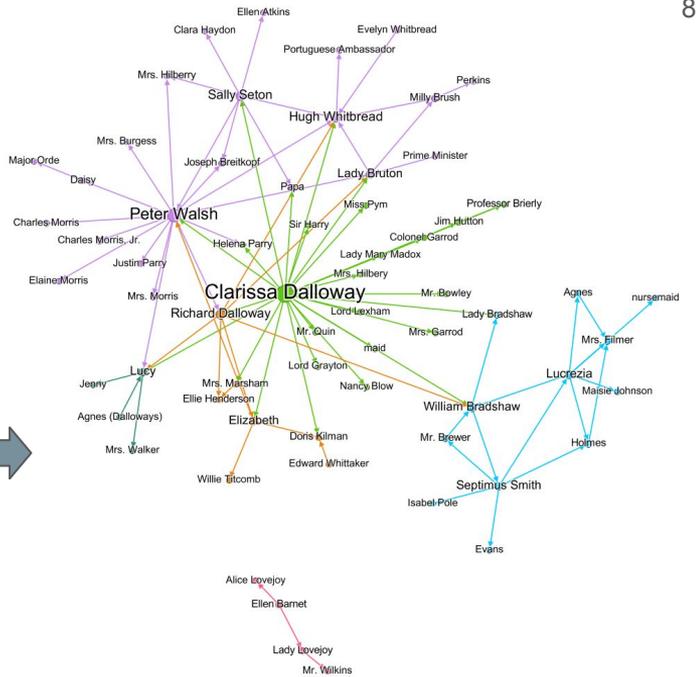
pw-cj



Transcribe to Excel



Network software (Gephi/NodeXL/etc.)



Id	Label	Betweenness Centrality
Clarissa Dalloway	Clarissa Dalloway	1126.0
Peter Walsh	Peter Walsh	543.75
William Bradshaw	William Bradshaw	503.166667
Lucrezia	Lucrezia	251.166667
Richard Dalloway	Richard Dalloway	192.916667
Hugh Whitbread	Hugh Whitbread	178.75
Lucy	Lucy	174.0
Sally Seton	Sally Seton	156.166667
Septimus Smith	Septimus Smith	146.166667
Lady Bruton	Lady Bruton	109.0

**Network Overview**

- Average Degree: 1.385 Run
- Avg. Weighted Degree: 1.385 Run
- Network Diameter: 6 Run
- Graph Density: 0.043 Run
- HITS: Run
- Modularity: 0.536 Run
- PageRank: Run
- Connected Components: 2 Run

**Node Overview**

- Avg. Clustering Coefficient: 0.476 Run
- Eigenvector Centrality: Run

**Edge Overview**

- Avg. Path Length: 2.986 Run

# Alternative annotation methods

- dumb digital annotation software
  - manual annotation of entities and relationships in software
  - more shareable, easily updated
  - still a lot of work
  - no network extraction/analysis
- automated annotation pipeline
  - automatic annotation of entities and relationships
  - no human input required
  - not very accurate
- semi-automated annotation pipeline
  - initial automatic annotation of entities and relationships
  - can be modified manually (high accuracy)
    - automatic suggestions aid in this process
  - more work than fully automated
  - less work than digital annotation software

Our proposed  
work

# Dumb digital annotation software

# Callisto

## Purpose

annotate text with entity information

## Input

text with or without entity mention annotations

## Output

entities, entity mentions, relationships

## Backend

Java

## Website

<http://mitre.github.io/callisto/>

Callisto - VOA19980414.1700.3225.sgm.tmx.rdc.apf.xml.txt.aif.xml

File Edit Format Tools Help

VOA19980414.1700.3225  
NEWS STORY

inter Milan defeated Spartak Moscow of Russia two to one and Lazio of Rome has played Athletic Madrid of Spain to a scoreless draw to qualify for an all Italian Nueva cup final next month. Lazio relied on defense, defense to reach their first ever European football final in a tense and scrappy match that was played at Rome's Olympic stadium. The draw gave the Romans a one nothing aggregate victory after its first leg one nothing road win. Meanwhile, Brazilian football star Renaldo scored both goals in inter Milan's two to one victory over Spartak Moscow to lead the team to the finals for the second year in a row. The second leg win allowed inter to advance on a four to two aggregate score after winning the fist leg match in Milan two weeks ago by an identical two to one score.

Mentions Entities Relationships Realties Time Ranges

ID	Offset	Mention	Head	Lexical Form	Role	Metonymy
	504	Brazilian	Brazilian	Name	GPE	null
	788	Milan	Milan	Name	LOC	null
	44	inter Milan	inter Milan	Name	null	null
	558	inter Milan	inter Milan	Name	null	null
	701	inter	inter	Name	null	null
	624	the team	team	Nominal	null	null
	105	Lazio of Rome	Lazio	Name	null	null

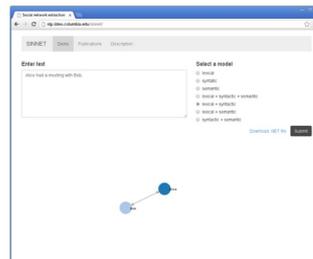
Text Swipe by Word | Font: 12pt. Default | Charset: UTF-8 | Task: RDC Task

# Automated annotation pipeline

# Past work on fully automated annotation pipelines

- End-to-end

- **input:** unstructured, unannotated text
- **output:** social network



SINNET  
Agarwal et al. (IJCNLP  
2013 Demo)

- First half of pipeline

- **input:** unstructured, unannotated text
- **output:** tokenized text with named entities resolved

Bamman, Underwood,  
and Smith (ACL 2014)

Agarwal et al. (IJCNLP  
2013 Demo)

- Second half of pipeline

- **input:** text, resolved entities, optional: mention annotations
- **output:** social network

Agarwal et al. (2010, 2013,  
2014, 2016)

# Automated annotation pipeline: End-to-end pipelines

# End-to-end systems: ad hoc

- windows / proximity
  - $(a,b)$  iff  $a$  and  $b$  mentioned within  $n$  words of each other  $k$  or more times
  - $n = 10, k = 3$ : Sack (Workshop on Computational Models of Narrative, 2013)
  - $n = 30, k = 1$ : Feild & Alexander (unpublished)
- speech
  - quoted speech attribution to extract  $(a,b)$  Elson & McKeown (2010)
- ...

# End-to-end system: SINNET

## Purpose

extract social network from text

## Input

text with or without entity mention annotations

## Output

visual network graph, export to gml and .net

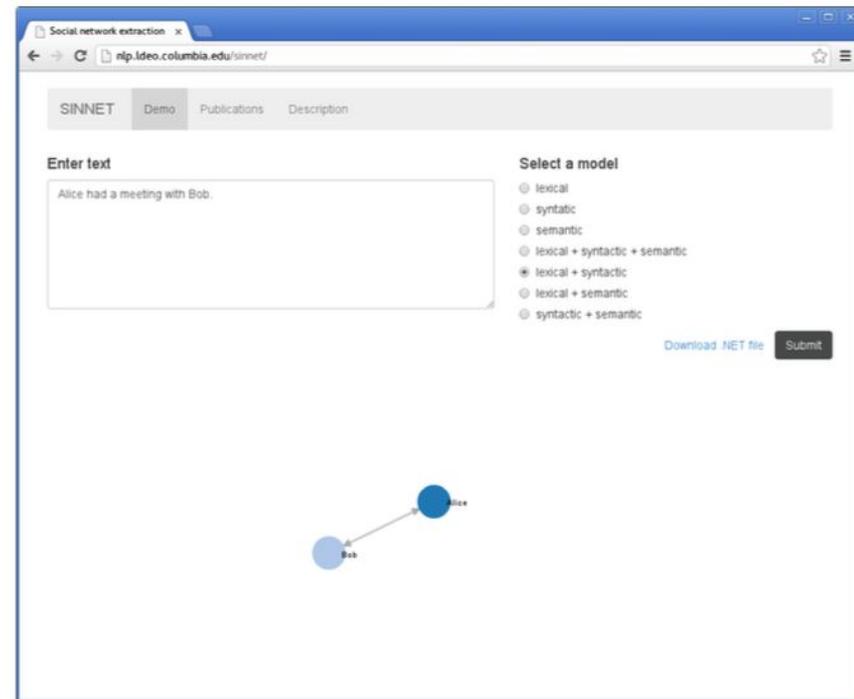
## Backend

Java

Uses Jet (Grishman, ACE Eval Workshop 2005) for named entity resolution, etc.

## Evaluation task:

- extract unweighted interactions
- excerpts from four 19th century novels
- Mean F1: **0.61**



SINNET (screenshot from Agarwal et al. IJCNLP 2013 Demo)

# Automated annotation pipeline: First half of pipeline

# BookNLP: Detect and resolve named entities

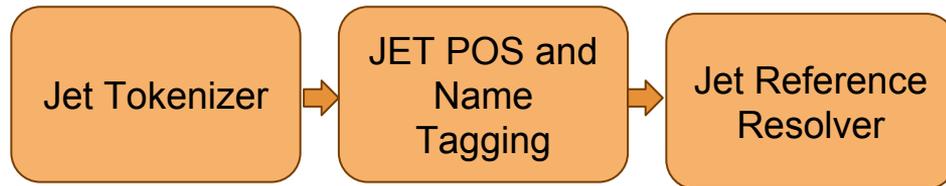
(Bamman, Underwood, and Smith, ACL 2014)



82.7% accuracy using 10-fold CV over a 3-book sample (*Pride and Prejudice*, *The Turn of the Screw*, *Heart of Darkness*)

# SINNET

(Agarwal, Kotalwar, Zheng, and Rambow, IJCNLP 2013 Demo)

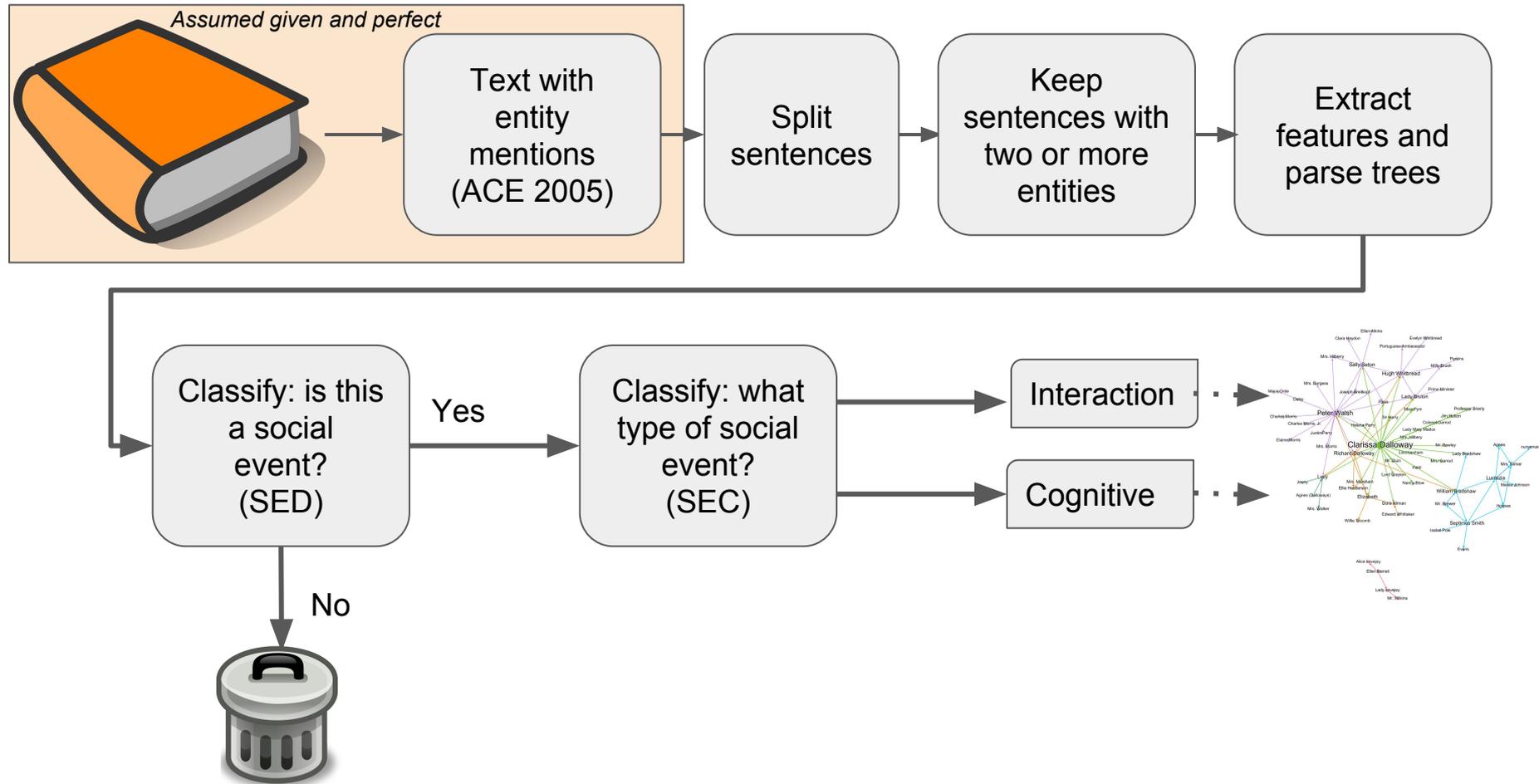


Jet (Grishman, ACE Eval Workshop 2005)

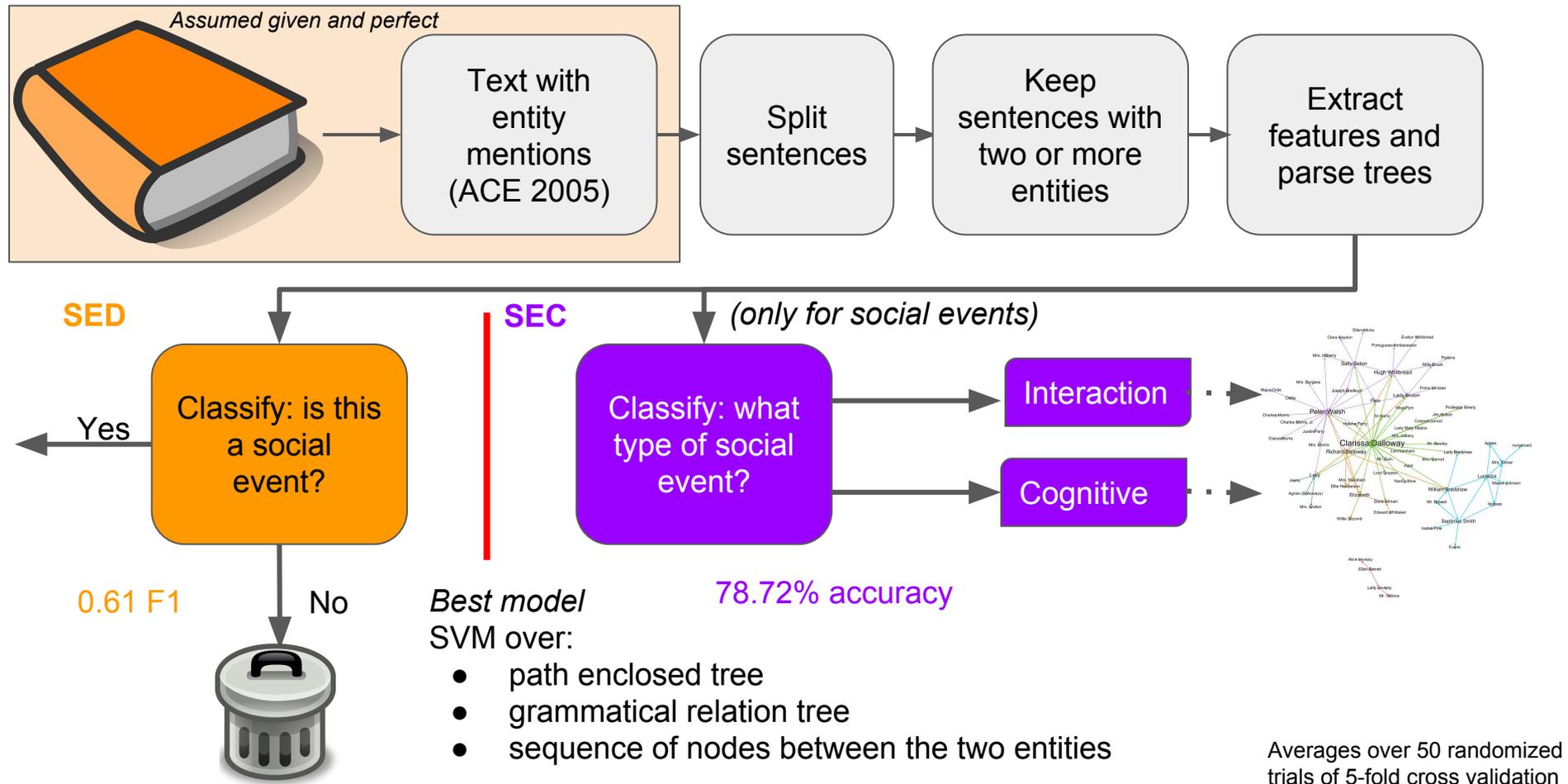
<https://cs.nyu.edu/grishman/jet/license.html>

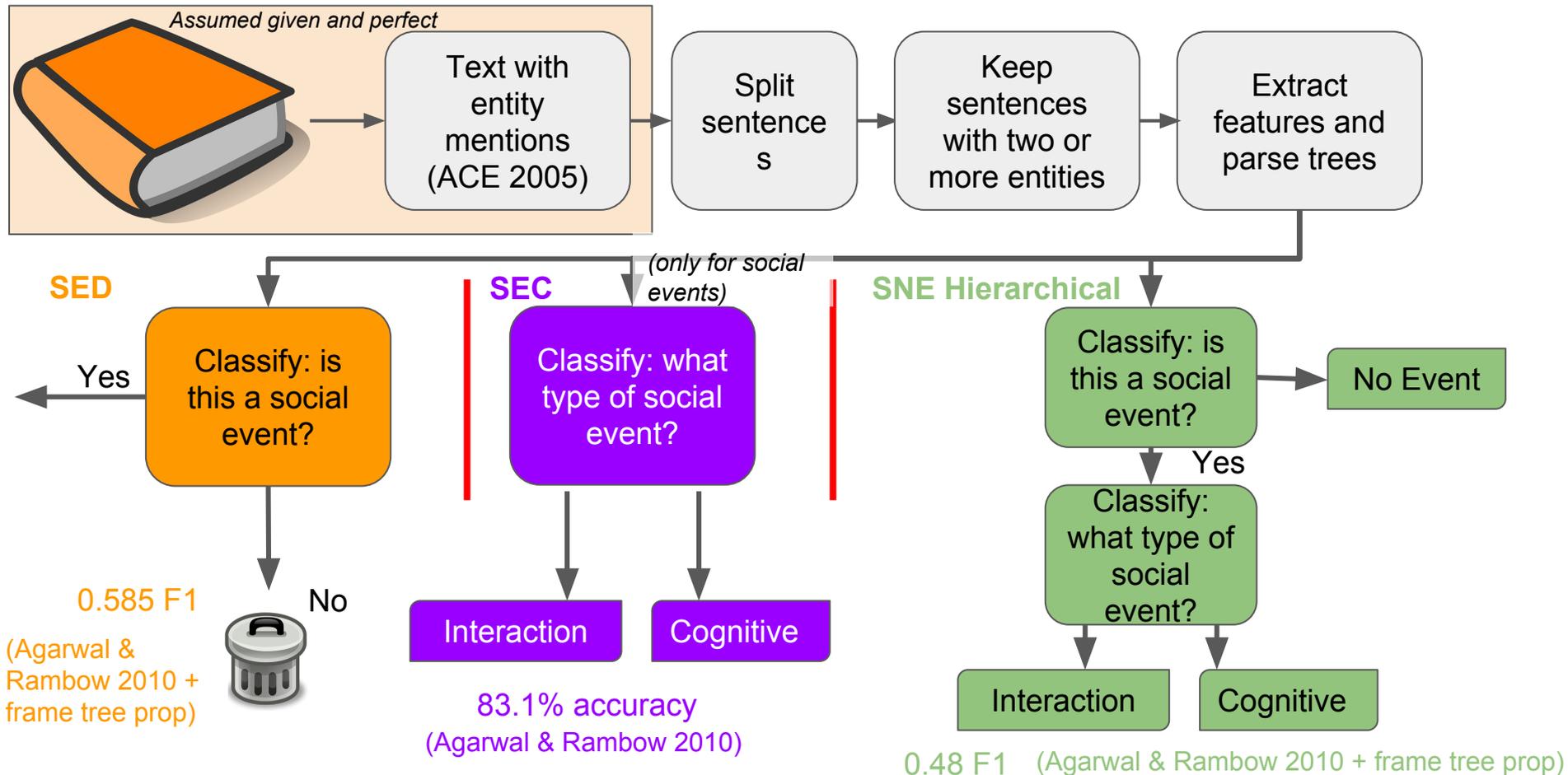
Automated annotation pipeline:  
Second half of pipeline

# Agarwal and Rambow (EMNLP 2010)

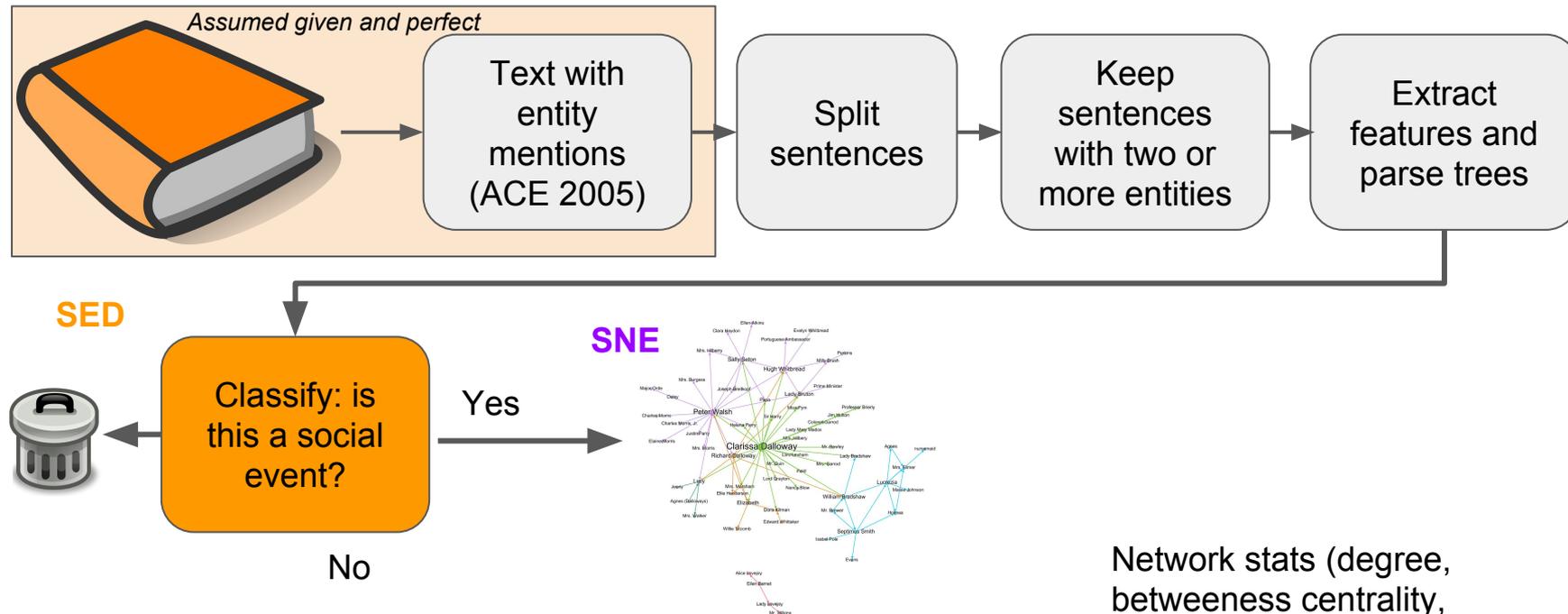


# Agarwal and Rambow (EMNLP 2010)





# Agarwal, Kotalwar, & Rambow (ICJNLP 2013)



Network stats (degree, betweenness centrality, connected components, density, etc.) are substantially different for baselines than the parse-tree based SVM.

	SED (F1)	Un-weighted SNE (F1)
B-Simple	0.56	0.57
BOW	0.48	0.61
<b>Parse tree-based SVM</b>	<b>0.61</b>	<b>0.68</b>

# Semi-automated annotation pipeline

# Goals

- easy to share and modify annotations
  - Google Docs for named entity and relationship annotations
  - common reference points
- less work than annotating from scratch
  - seed with automatic methods
  - generate recommendations based on automatic methods + learning from annotations
- high-confidence results
  - Agarwal (2016) found F1 scores in the range 0.51–0.70 across four novels
  - scholars need more certainty of the accuracy of the annotations and resulting network

# Our vision

# Characters

## Mrs. Dalloway

Mrs. Dalloway  
Clarissa

## Peter Walsh

The elderly man  
Peter Walsh

## Lucy

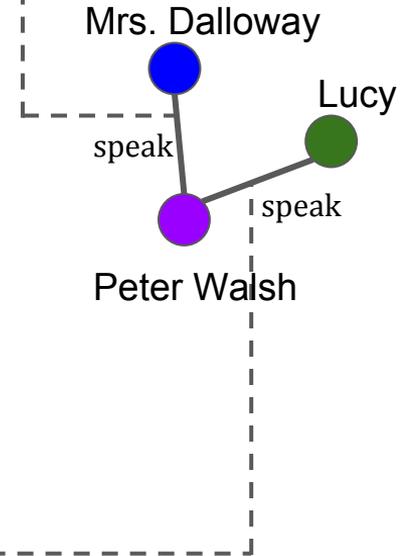
Lucy

"Mrs. Dalloway will see me," said the elderly man in the hall. "Oh yes, she will see me," he repeated, putting Lucy aside very benevolently, and running — — — upstairs ever so quickly. "Yes, yes, yes," he muttered as he ran upstairs. "She will see me. After five years in India, Clarissa will see me."

"Who can—what can," asked Mrs. Dalloway (thinking it was outrageous of the day she was giving a party), hearing a step on the stairs. She heard a hand upon the door. She made to hide her dress, like a virgin protecting chastity, respecting privacy. Now the brass knob slipped. Now the door opened, and in came—for a single second she could not remember what he was called! so surprised she was to see him, so glad, so shy, so utterly taken aback to have Peter Walsh come to her unexpectedly in the morning! (She had not read his letter.)

"And how are you?" said Peter Walsh, positively trembling; taking both her hands; kissing both her hands. She's grown older, he thought, sitting down. I shan't tell her anything about it, he thought, for she's grown older. She's looking at me, he thought, a sudden embarrassment coming over him, though he had kissed her hands. Putting his hand into his pocket, he took out a large pocket-knife and half opened the blade.

# Relationships



# Characters

# Relationships

"Mrs. Dalloway will see me," said the elderly man in the hall. "Oh yes, she will see me, putting Lucy aside very benevolently, and he ran upstairs. "She will see me. After five years in India, Clarissa will see me."

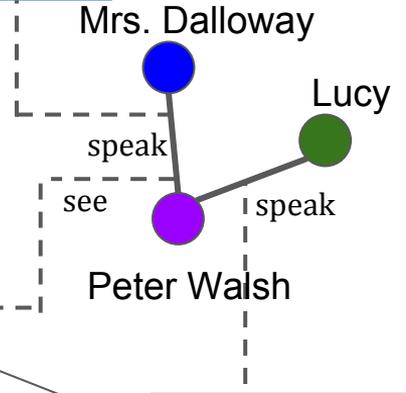
the elderly man

Peter Walsh  
Mrs. Dalloway  
Lucy  
More

Re-assign mentions to other entities

- Mrs. Dalloway  
Mrs. Dalloway  
Clarissa
- Peter Walsh  
The elderly man  
Peter Walsh
- Lucy  
Lucy

"Who can—what can," asked Mrs. Dalloway (thinking it was outrageous of the day she was giving a party), hearing a step on the stairs. She heard a hand upon the door. She made to hide her dress, like a virgin protecting chastity, respecting privacy. Now the brass knob slipped. Now the door opened, and in came—for a single second she could not remember what he was called! so surprised she was to see him, so glad, so shy, so utterly taken aback to have Peter Walsh come to her unexpectedly in the morning! (She had not read his letter.)



Group/split/rename character lists

"And how are you?" said Peter Walsh, positively trembling; taking both her hands; kissing both her hands. She's grown older, I shan't tell her anything, she's grown older. She's looking embarrassed coming in, she kissed her hands. Putting down the letter, she took out a large pocket-knife and half opened the blade.

Peter Walsh  
Mrs. Dalloway  
Lucy  
More

Create new or modify existing relationships

Select new mentions and tie them to an entity

https://www.isittoasted.com/app?mode=analyze&cohort=doctors&doctype=copy&c

**toasted** FAQs Ethics Examples ▾ Sign up Login

I'm writing **copy** ▾ for **doctors** ▾

Doctors Without Borders-Independent, **impartial**, **neutral**

Instead, try ...

It's easy to write inspiring words that do **non-partisan** mission. It's much harder to put those **principles** **into practice**. Since **neutral** Doctors Without Borders / Médecins Sans Frontières (MSF) has operated with a **equitable** **commitment** to independence, **impartiality**, and **neutrality**.

These **principles** are what make **even-handed** bond rapidly to **emergencies** and provide lifesaving medical care in situations where many other organizations can't or won't. This **commitment** ensures that we can help people most in need, rather than following other **priorities** — whether they be those of politicians, funders, or the media.

The **impact** and **efficacy** of our ability to put these **principles** **into practice** **was recognized** in 1999 when MSF **was awarded** the Nobel Peace Prize.

Where we go

Write  Analyze

You have pending changes. Please press 'Toast' to get an updated score.

**fairly effective** ⓘ

Your writing style is **48% better than average** for an audience of doctors.

- The readability is at a tenth grade level (a little too low). Use more complex sentences.
- Consider using less of the passive voice.
- Click on the colored terms for suggestions. The darker the blue, the better the choice.

© Toasted • [Terms of Use](#) • [Contact Us](#) • [Careers](#)

[www.isittoasted.com](http://www.isittoasted.com)

# Automatic assist

- automatic propagation of changes
  - changing the entity associated with a mention changes everything that mention is used in
    - relationships
    - character lists
- several ranking problems
  - entity suggestion when merging entity aliases
  - entity suggestion when highlighting a mention
  - relationship suggestions when selecting two mentions to relate in the text
- we can get better at recommendations by observing annotator behavior
  - the system should get better over time

Mrs. Dalloway  
Mrs. Dalloway  
Clarissa

Peter Walsh  
Peter Walsh

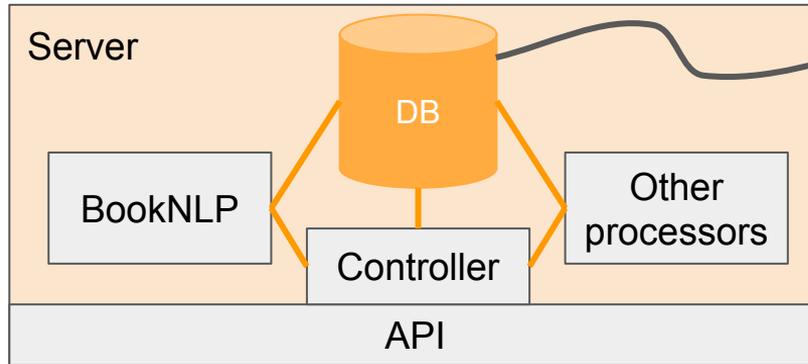
Lucy  
Lucy

The elderly man  
The elderly man

Merge with...  
Peter Walsh  
Lucy  
Mrs. Dalloway  
more

# Feature wish list

- public and private
  - book uploads
  - annotations
- annotations can be shared with other users
  - read-only / read+write permissions can be granted to specific users or the public
- multiple book uploads are only processed once
  - using checksums, if a book has been uploaded and processed in the past, it's not processed again unless explicitly told to do so
  - this should make the system faster with heavy load (e.g., in a classroom setting)



## Data

- users
- permissions
- processed texts
- annotations of a text
  - entities / groups
  - entity locations
  - relationships

Upload/browse  
books

Create/select  
annotations

Share  
annotations

*Manager view*

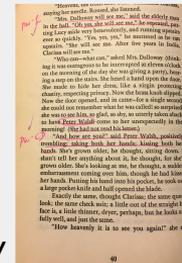
Browser

manage  
entities

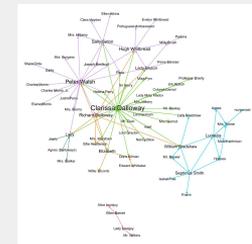


*Annotation view*

manage  
annotations



visualize  
relationships



The screenshot shows a web browser window with the following elements:

- Browser Tab:** "Character extraction" with a close button (X) and a plus sign (+) for a new tab.
- Address Bar:** "localhost:3535/character-list.php" with navigation icons (back, forward, refresh, home) and utility icons (info, shield, star, download, printer, search, etc.).
- Page Header:** "When Entities Interact" | Texts | Annotations | Upload text | test ▾
- Main Content:**
  - Your texts**
    - No texts found :( [Upload one!](#)
  - All texts**
    - [Dracula](#) (processed: yes)
    - [Huckleberry Finn](#) (processed: yes)
    - [Dorian Gray](#) (processed: yes)
    - [Scarlet Letter](#) (processed: no)
  - [Refresh](#) button

Character extraction

localhost:3535/character-list.php#annotations

When Entities Interact   Texts   Annotations   Upload text   test ▾

## Extracted characters

Fork annotation   List annotations

Group selected

- Samuel F. Billington  Hamlet  Berkeley Hotel  Billington
- Quincey P. Morris  QUINCEY P. MORRIS
- Mitchell , Sons , & Candy  Mitchell  Mitchell , Sons
- Enoch  Carpathians
- Mina Harker  Customs  MINA HARKER
- Arthur  Arthur Holmwood  ARTHUR  Holmwood
- Lloyd
- Dear Madam
- Madam Mina
- Immanuel Hildesheim
- Shakespeare
- Doolittle
- Attila
- King  Cross  King 's Cross
- Mr. Billington
- Lor
- Journal
- JONATHAN HARKER  Harker  Jonathan  Jonathan Harker
- Gott
- Catherine
- Carter  Carter Paterson

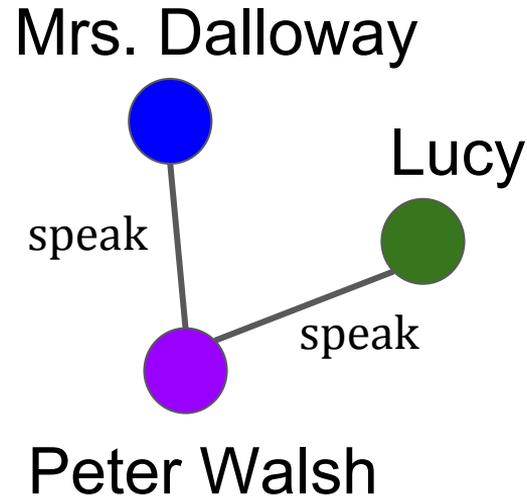
Users can modify automatically extracted entity lists:

- group
- split
- delete
- add

Users can also fork annotations.

# Moving forward

- complete the prototype system to allow for all the annotations
  - mentions
  - relationships
- add in client-side graphing (d3, probably)
- add in sharing
- add advanced recommendations
- test!
- rinse and repeat:
  - update
  - test with some real users



# Summary

- automatic methods of annotation are scalable
  - but lack the accuracy and adaptability of manual methods
- manual methods are (generally) accurate and flexible
  - but lack scalability and sharability of digital methods
- a semi-automated method may be in the Goldilocks zone
  - also provide a framework for getting "free" training data
  - introduce lots of new subproblems

# Thanks

- Phil Lombardo—my direct collaborator
- Sam Alexander—inspiration for the project, insights from the lit world
- Niranjan Balasubramanian—for early advice on some of the related NLP
- Laura Dietz—my NLP connection and UNH host
- all of you for listening!

# Questions?