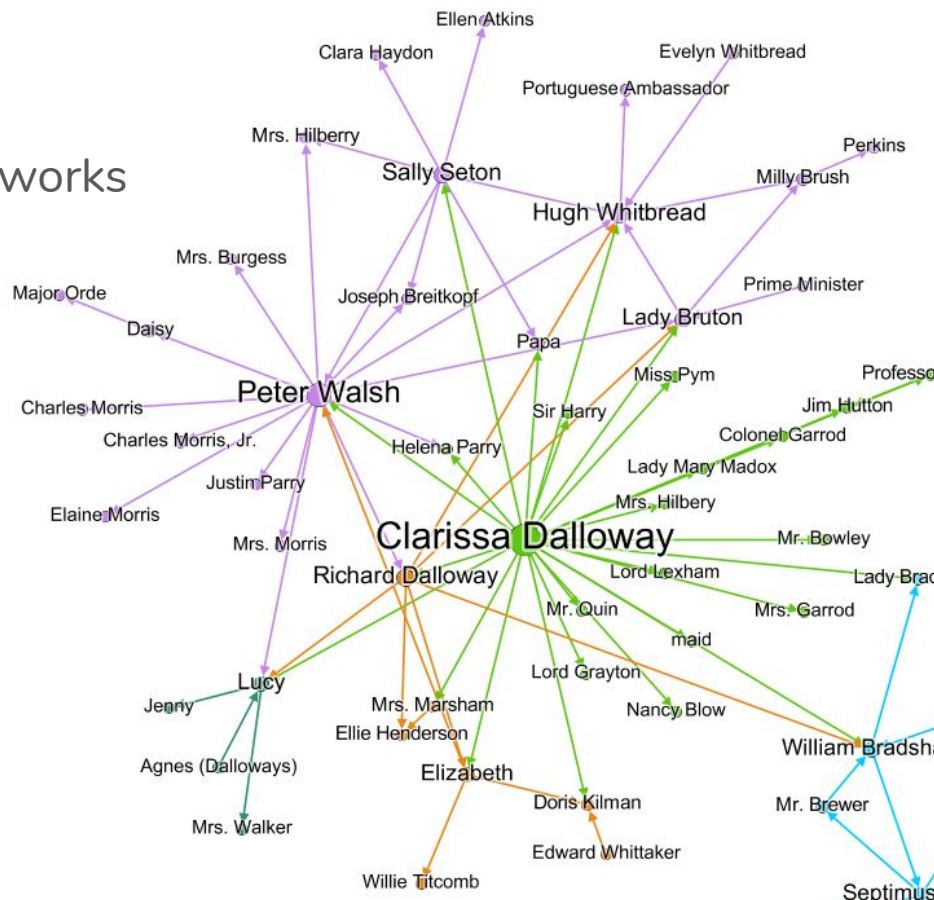


Developing EntiTies: Sabbatical Overview

Hank Feild
Endicott College
March 11, 2020

Outline

- Character networks
 - Methods of establishing character networks
 - EntiTies
 - Where we are & future directions
- Major Orde



A page from Mrs. Dalloway by Virginia Woolf

"Heavens, the front-door bell!" exclaimed 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 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.

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

Lucy

speak

speak

Peter Walsh

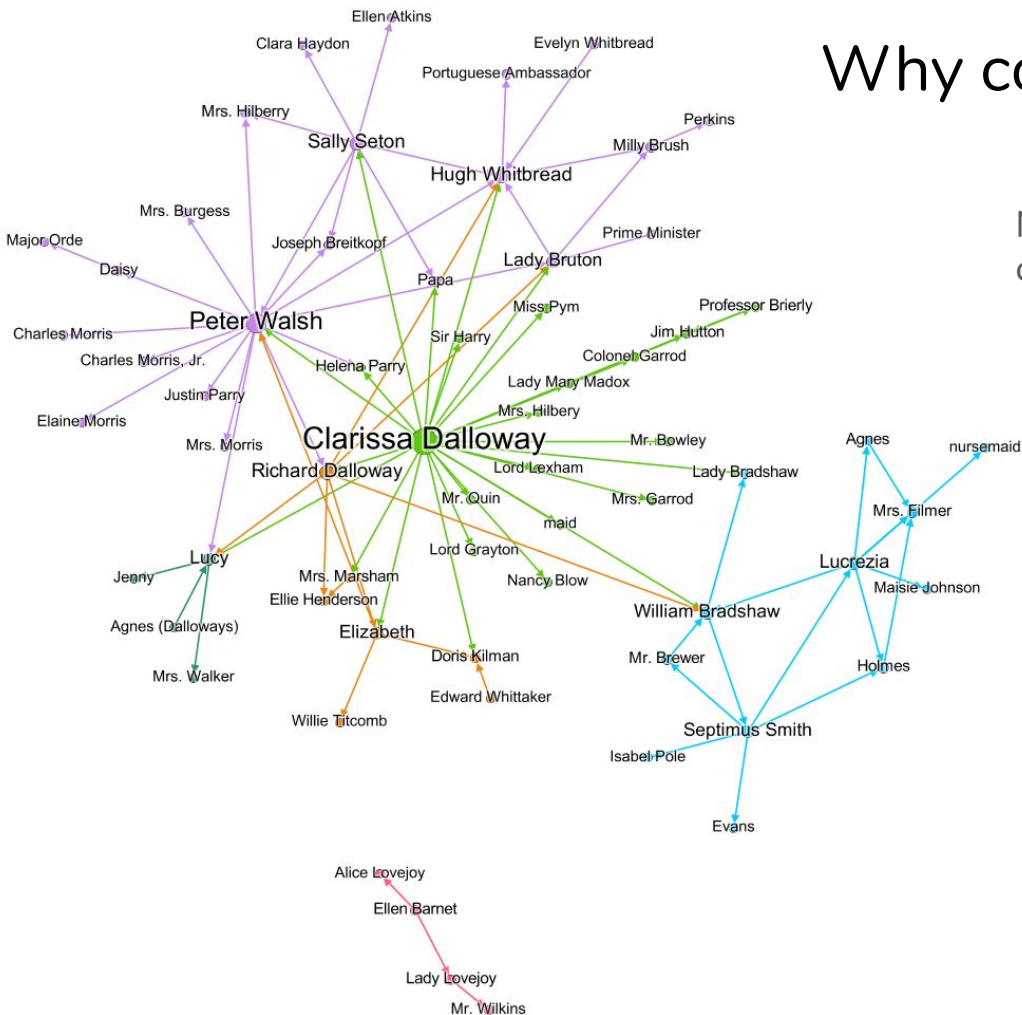
Nodes are characters

Edges or ties are relationships

Why consider character networks?

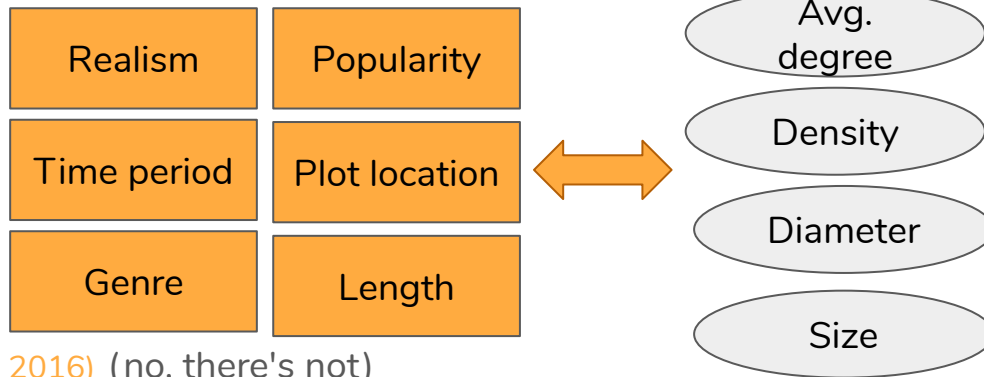
Network analysis gives us many ways to quantify characteristics of a network:

- the number of connections each node has
- how important each node is for bridging disparate parts of the network
- the communities and clusters within the network
- the number of connected components
- the average number of hops to get between any two nodes
- network diameter (widest distance from one end to the other)



Literary questions we might answer using networks

- how do conversation networks differ between rural and urban novels in the 19th Century? (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)



Beyond character networks

Boston Novr. 13. 1779

My dearest Friend

I have just sent Mr. Thaxter, Johnny and Stephens with the Things on Board. I shall go with Charles at four O Clock. It is now three. Have seen the Captain, and the Navy Board &c.

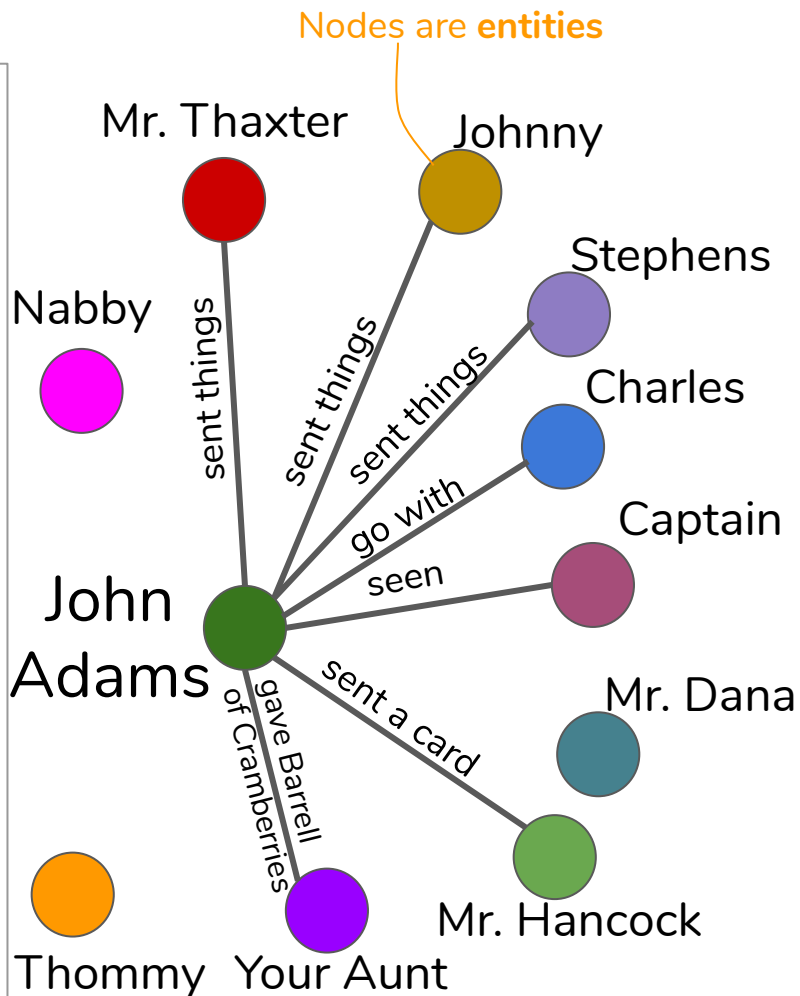
It is proposed to sail tomorrow. Perhaps however, it may not be till next day. Mr. Dana will come on board at Nine tomorrow.

Mr. Hancock has sent me a Card, to invite me to go on board with him in the Castle Barge [based at Castle Island in Boston Harbor, used for ceremonial purposes] . -- Dont make many Words of this.

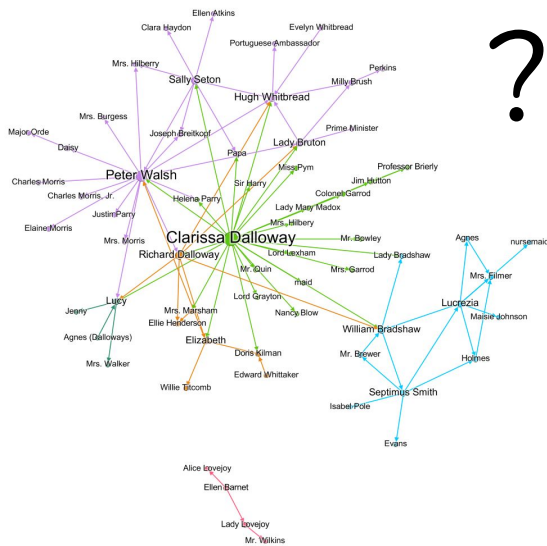
Your Aunt has given me a Barrell of Cramberries. I shall make a good Use of them, I hope.

Let me intreat you, to keep up your Spirits and throw off Cares as much as possible. Love to Nabby and Thommy. We shall yet be happy, I hope and pray, and I dont doubt it. I shall have Vexations enough, as usual. You will have Anxiety and Tenderness enough as usual. Pray strive, not to have too much. I will write, by every Opportunity I can get.

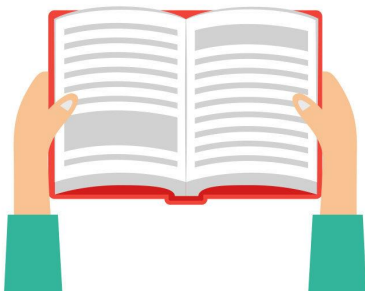
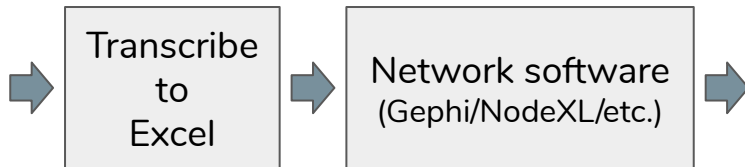
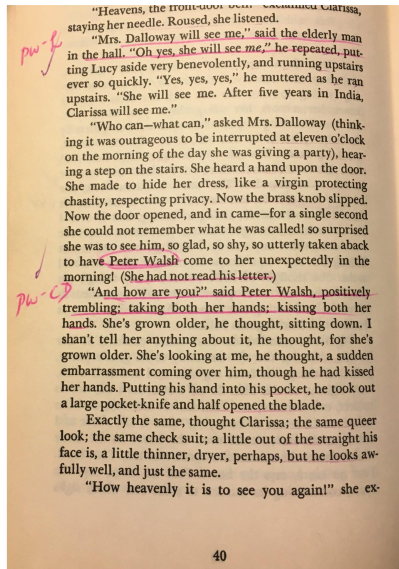
Yours, ever, ever yours,
John Adams



How do we get from to ?



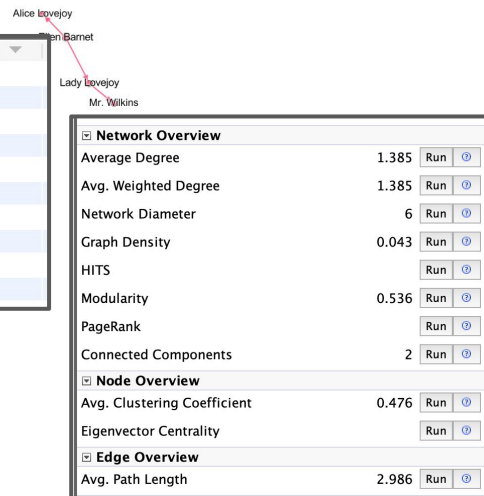
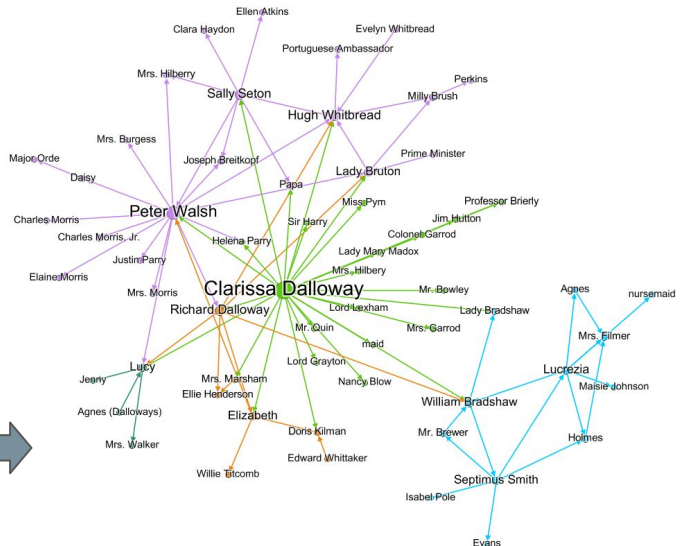
Manual annotation pipeline



Drawbacks

- difficult to **share** and **modify** annotated text
- tedious
- prone to errors during transcription

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



Fully automated pipeline

Title: Mrs. Dalloway
Author: Virginia Woolf

Mrs. Dalloway said she would buy the flowers herself.

For Lucy had her work cut out for her. The doors would be taken off their hinges; Rumpelstiltskin's men were coming. And then, thought Clarissa Dalloway, what a morning—fresh as if issued to children on a beach.

What a lark! What a plunge! For so it had always seemed to her, when, with a little squeak of the hinges, which she could hear now, she had burst open the French windows and plunged at Bourton into the open air. How fresh, how calm, stiller than this of course, the air was in the early morning; like the flap of a wave; the kiss of a wave; chill and sharp and yet (for a girl of eighteen as she then was) solemn, feeling as she did, standing there at the open window, that something awful was about to happen; looking at the flowers, at the trees with the smoke winding off them and the rooks rising, falling; standing and looking until Peter Walsh said, "Musing among the vegetables?"—was that it?—"I prefer men to cauliflowers"—was that it? He must have said it at breakfast one morning when she had gone out on to the terrace—Peter Walsh. He would be back from India one of these days, June or July, she forgot which, for his letters were awfully dull; it was his sayings one remembered; his eyes, his pocket-knife, his smile, his grumpiness and, when millions of things had utterly vanished—how strange it was!—a few sayings like this about cabbages.

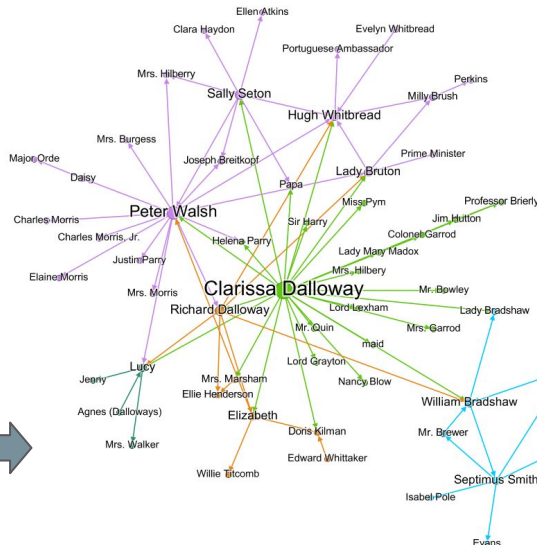
She stiffened a little on the kerb, waiting for Durtall's van to pass. A charming woman, Scrope Purvis thought her (knowing her as

List of character
s and
their
aliases

(not
necessarily
required)

Program that
finds mentions
of characters
and establishes
ties between
them

Network software
(Gephi/NodeXL/etc.)



Alice Lovejoy
Ellen Barnet
Lady Lovejoy
Mr. Wilkins

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

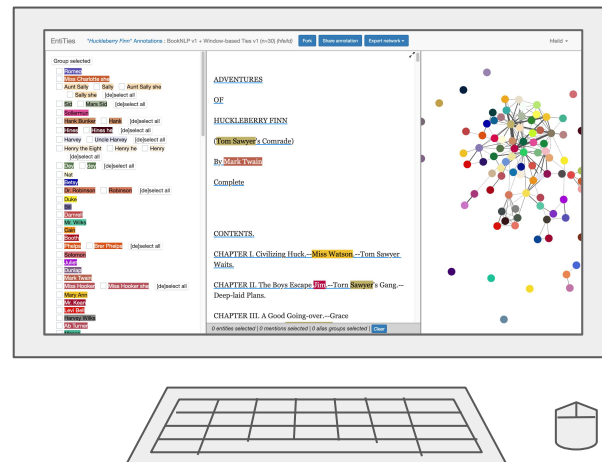
Drawbacks

- inaccurate (by quite a bit)
- difficult to **verify** or **modify** annotated text

How can we address these drawbacks?

An application that...

- keeps track of annotations
 - entity/mention locations
 - tie locations
- allows different workflows
 - manual annotation
 - automated annotation
 - semi-automated annotation
- allows (encourages!) sharing
 - and forking
- made with general end users in mind
 - literary scholars, historians, lawyers, etc.
 - not just linguists and computational linguists



EntiTies

<http://entities.greenbirdlabs.dev>

Group selected

ADVENTURES

OF

HUCKLEBERRY FINN

(Tom Sawyer's Comrade)

By Mark Twain

Complete

CONTENTS.

CHAPTER I. Civilizing Huck.--Miss Watson.--Tom Sawyer Waits.

CHAPTER II. The Boys Escape Jim.--Torn Sawyer's Gang.-- Deep-laid Plans.

CHAPTER III. A Good Going-over.--Grace

Group selected

- ☐ Romeo
- ☐ Miss Charlotte she
- ☐ Aunt Sally ☐ Sally ☐ Aunt Sally she
- ☐ Sally she [de]select all
- ☐ Sid ☐ Mars Sid [de]select all
- ☐ Sollemun
- ☐ Hank Bunker ☐ Hank [de]select all
- ☐ Hines ☐ Hines he [de]select all
- ☐ Harvey ☐ Uncle Harvey [de]select all
- ☐ Henry the Eight ☐ Henry he ☐ Henry
- [de]select all
- ☐ Day ☐ dey [de]select all
- ☐ Nat
- ☐ Betsy
- ☐ Dr. Robinson ☐ Robinson [de]select all
- ☐ Duke
- ☐ Bill
- ☐ Damrell
- ☐ Mr. Wilks
- ☐ Cain
- ☐ Booth
- ☐ Phelps ☐ Brer Phelps [de]select all
- ☐ Solomon
- ☐ Juliet
- ☐ Dunlap
- ☐ Mark Twain
- ☐ Miss Hooker ☐ Miss Hooker she [de]select all
- ☐ Mary Ann
- ☐ Mr. Kean
- ☐ Levi Bell
- ☐ Harvey Wilks
- ☐ Ab Turner
- ☐ Moses

ADVENTURES

OF

HUCKLEBERRY FINN

(Tom Sawyer's Comrade)

By Mark Twain

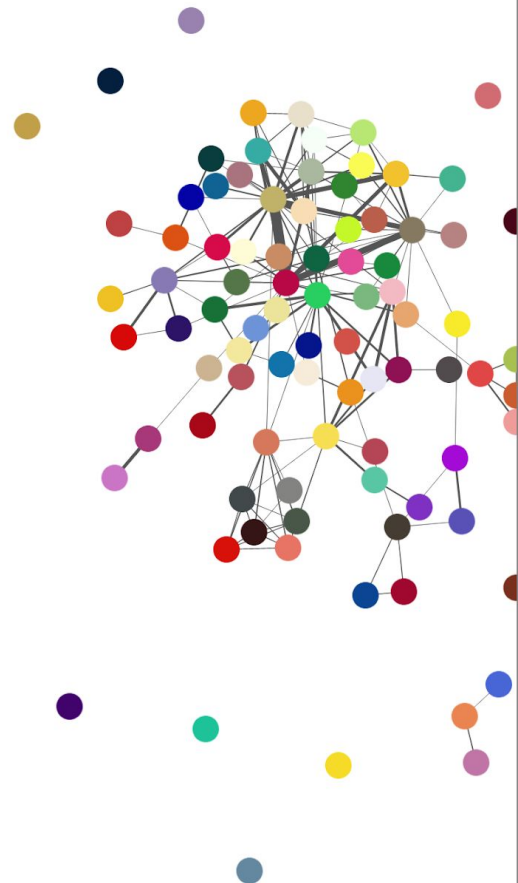
Complete

CONTENTS.

CHAPTER I. Civilizing Huck.--Miss Watson.--Tom Sawyer Waits.

CHAPTER II. The Boys Escape Jim.--Torn Sawyer's Gang.--Deep-laid Plans.

CHAPTER III. A Good Going-over.--Grace



EntiTies as a research framework

- in the literature, there are several techniques to
 - identify entities and their mentions in text
 - extract ties between entities in a segment of text
- which work best? and for which use cases?
- EntiTies has a built in user study feature
 - a study consists of tasks
 - links to external resources, e.g., a Qualtrics questionnaire
 - annotation tasks
 - tasks can be grouped into orderings
 - interactions with the annotation interface are logged for analysis

	Group 1	Group 2	Group 3	Group 4
	<i>Pre-annotation questionnaire</i>			
Task 1	Excerpt 0 Condition A	Excerpt 0 Condition B	Excerpt 0 Condition B	Excerpt 0 Condition A
Task 2	Excerpt 1 Condition A	Excerpt 2 Condition B	Excerpt 1, Condition B	Excerpt 2, Condition A
Task 3	Excerpt 0 Condition B	Excerpt 0 Condition A	Excerpt 0 Condition A	Excerpt 0 Condition B
Task 4	Excerpt 2 Condition B	Excerpt 1 Condition A	Excerpt 2, Condition A	Excerpt 1, Condition B
	<i>Post-annotation questionnaire</i>			

EntiTies sp2020_40_697 ▾

EntiTies: ENG235 Twentieth Century Fiction (SP20)

Please complete each of the following tasks in the order listed. Refrain from multi-tasking while working on each task.

Tasks

1. Pre-annotation questionnaire ✓
2. Annotate "Their eye were watching God" Ch. 1-2 from scratch
3. Annotate "Their eye were watching God" Ch. 19-20 from scratch
4. Correct "Their eye were watching God" Ch 1-2 annotation
5. Correct "Their eye were watching God" Ch 10-11 annotation
6. Post-annotation questionnaire

Research & Development

- version 1: Demonstration paper at CHIIR 2020
- pilot study (Sam's Twentieth Century Fiction class); n=21
 - test the interface
 - initial evaluation of workflows: manual vs. semi-automated vs. fully automated
 - test the study system (interface, data collection & analysis)
- this summer
 - improve interface
 - reduce clicking + scrolling, iron out glitches
 - more network viz options
 - smart recommendations when you highlight or click on a mention or tie
 - implement and improve upon algorithms for:
 - entity identification (targeted to other text types: historical, news, legal, etc.)
 - link extraction (Agarwal et al. 2010–2016)

Future directions

- user studies
 - students using it for class activities
 - scholars using it for their research
- recruit instructors use EntiTies in their classes
 - let me know if you're interested!
- recruit scholars to use it in their research
 - could you use EntiTies for your research?
- apply for NEH or NSF grants
 - targeted towards further developing and using EntiTies

Contributors



Tim Amello
Computer Science Major
('21)



Phil Lombardo



Sam Alexander