Exploring and Visualizing Patterns in Literary Text Collections with FeatureLens

Anthony Don, Catherine Plaisant, Tanya Clement
Elena Zheleva, Machon Gregory, Sureyya Tarka,
Celeste Paul, Ben Shneiderman

and others from the MONK project, in particular Loretta Auvil and Matt Kirschenbaum
Exploring and Visualizing Patterns in Literary Text Collections with FeatureLens

Anthony Don, Catherine Plaisant, Tanya Clement
Elena Zheleva, Machon Gregory, Sureyya Tarka, Celeste Paul, Ben Shneiderman

and others from the MONK project, in particular Loretta Auvil and Matt Kirschenbaum
Supporting Literary Scholars

The MONK Project
Metadata Offer New Knowledge

Humanities scholars, computer scientists and information specialists

University of Illinois, Urbana-Champaign
School of Library and Information Sciences
National Center for Supercomputing Applications

Northwestern University
Academic Technologies

University of Maryland
Human-Computer Interaction Laboratory
Maryland Institute for Technology in the Humanities

University of Alberta
Department of English

MacMaster University
Humanities Visualization

University of Nebraska, Lincoln
Center for Digital Research in the Humanities

Guiding Principles
Web-deliverable services
Visualizations as interface
State of the art text mining

18th and 19th Century
British and American Literature

Follow up of Nora Project
Presentation last year:
Exploring the erotics in Emily Dickinson’ letters with text mining and visual interfaces
Motivating problem

- Thesis of Tanya Clement, PhD student from English Department
- Gertrude Stein’s “The Making of Americans” (MoA)
- Study the use of repetition

<table>
<thead>
<tr>
<th></th>
<th>Uncle Tom’s Cabin</th>
<th>Moby Dick</th>
<th>The Making of Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # words</td>
<td>190,906</td>
<td>220,254</td>
<td>517,207</td>
</tr>
<tr>
<td>Unique words</td>
<td>11,730</td>
<td>14,512</td>
<td>5,329</td>
</tr>
</tbody>
</table>
Sample paragraph

• [1086] Always from the beginning there was to me all living as repeating. This is now a description of my feeling. As I was saying listening to repeating is often irritating, always repeating is all of living, everything in a being is always repeating, more and more listening to repeating gives to me completed understanding.
Examples of “Literary” questions about MoA

How does linguistic and syntactic variation reflect, advance, or deter from the progression of the text’s more abstract features such as argument and narrative?

Do the changes in repetition correspond to the novel’s evolving theories about identity and representation? And how?
Sample paragraph

• [1086] Always from the beginning there was to me all living as repeating. This is now a description of my feeling. As I was saying listening to repeating is often irritating, always repeating is all of living, everything in a being is always repeating, more and more listening to repeating gives to me completed understanding.
[1086] Always from the beginning there was to me all living as repeating. This is now a description of my feeling. As I was saying listening to repeating is often irritating, always repeating is all of living, everything in a being is always repeating, more and more listening to repeating gives to me completed understanding.
Words

• [1086] Always from the beginning there was to me all living as repeating. This is now a description of my feeling. As I was saying listening to repeating is often irritating, always repeating is all of living, everything in a being is always repeating, more and more listening to repeating gives to me completed understanding.
Higher level concept

[1086] Always from the beginning there was to me all living as repeating. This is now a description of my feeling. As I was saying listening to repeating is often irritating, always repeating is all of living, everything in a being is always repeating, more and more listening to repeating gives to me completed understanding.
Ngrams (sets of consecutive words)

• [1086] Always from the beginning there was to me all living as repeating. This is now a description of my feeling. As I was saying listening to repeating is often irritating, always repeating is all of living, everything in a being is always repeating, more and more listening to repeating gives to me completed understanding.
Questions we try to address

• **What text features are highly repeated in the text?**
  – Frequent words
  – Frequent n-grams (consecutive words)
  – Frequent patterns of n-grams (more “fuzzy” matches)

• **How do they change over time (i.e. along the text)?**
  – Locate features in text (overview + reading)
  – Compare features
  – Study the shape of change
  – Find features that exhibit specific shapes (e.g. spike)
Trying existing tools – Text Arc

Trying existing tools – Arc Diagrams

Arc Diagram of Für Elise

Trying existing tools – Arc Diagrams

Arc Diagram of one page of MoA

Trying existing tools

SpotFire

TimeSearcher
The 2007 State of the Union Address

Over the years, President Bush’s State of the Union Address has averaged almost 5,000 words each, meaning the President has delivered over 34,000 words. Some words appear frequently while others appear only sporadically. Use the tools below to analyze what Mr. Bush has said.

Use of the phrase "Iraq" in past State of the Union Addresses

<table>
<thead>
<tr>
<th>Year</th>
<th>2001*</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iraq</td>
<td>0</td>
<td>2</td>
<td>21</td>
<td>24</td>
<td>27</td>
<td>10</td>
<td>34</td>
</tr>
</tbody>
</table>

Compared with other words

- **Iraq**
  - 2 21 24 27 15 34
- **Afghanistan**
  - 13 3 5 3 2 4
- **Economy(ic)**
  - 6 7 13 17 14 23 6
- **Insurance**
  - 2 - 1 6 1 3 14
- **Iraq/Iraqi(s)**
  - 2 21 24 27 15 34
- **Iran**
  - 2 3 1 3 6 6
- **Oil**
  - 1 - - - 3 9
- **Social Security**
  - 15 2 2 2 18 3 2

* As a newly elected president, Mr. Bush did not deliver a formal State of the Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

---

The word in context

Iraq continues to flaunt its hostility toward America and to support terror. The Iraqi regime has plotted to develop anthrax, and nerve gas, and nuclear weapons for over a decade. This is a regime that has already used poison gas to murder thousands of its own citizens — leaving the bodies of mothers huddled over their dead children. This is a regime that agreed to international inspectors — then kicked out the inspectors. This is a regime that has something to hide from the civilized world.

---

*citizen* — [NYTimes](http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html)
was almost completely clearly feeling being one being living. He was almost completely clearly feeling in being one being living. He was one almost completely clearly feeling about needing being one being living, about not needing being one being living, about knowing being one being living, about being certain of being one being living, about not being certain of being one being living. He was almost completely clearly feeling something, he was almost completely clearly feeling everything of any such thing.

2948

He was almost completely clearly feeling that someone was different from any other one. He was almost completely clearly feeling this thing that was not this thing that was different from any other one. He was mentioning being certain of this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was almost completely clearly feeling something about this thing about
THE PRESIDENT: Mr. Speaker, Mr. Vice President, members of Congress: It's a great privilege to be here to outline a new budget and a new approach for governing our great country. I thank you for your invitation to speak here tonight. I know Congress had to formally invite me, and it could have been a close vote. (Laughter.) So, Mr. Vice President, I appreciate you being here to break the tie. (Laughter.)

I want to thank so many of you who have accepted my invitation to come to the White House to discuss important issues. We're off to a good start. I will continue to meet with you and ask for your input. You have been kind and candid, and I thank you for making a new President feel welcome. (Applause.)

The last time I visited the Capitol, I came to take an oath on the steps of this building. I pledged to honor our Constitution and laws. (Applause.) And I asked you to join me in setting a tone of civility and respect in Washington. (Applause.)

I hope America is noticing the difference, because we're making progress. Together, we are...
Americans who have lost their jobs need our help and I support extending unemployment benefits and direct assistance for health care coverage. (Applause.) Yet, American workers want more than unemployment checks -- they want a steady paycheck. (Applause.)

When America works, America prospers, so my economic security plan can be summed up in one word: jobs. (Applause.)

Good jobs begin with good schools, and here we've made a fine start. (Applause.) Republicans and Democrats worked together to achieve historic education reform so that no child is left behind. I was proud to work with members of both parties: Chairman John Boehner and Congressman George Miller.

(Applause.) Senator Jud Gregg.

(Applause.) And I was so proud of our work, I even had nice things to say about my friend, Ted Kennedy. (Laughter and applause.) I know the folks at the Crawford coffee shop couldn't believe I'd say such a thing -- (laughter) -- but our work on this bill shows what is possible if we set aside posturing and focus on results. (Applause.)
The last time I visited the Capitol, I came to take an oath on the steps of this building. I pledged to honor our Constitution and laws. (Applause.) And I asked you to join me in setting a tone of civility and respect in Washington. (Applause.)

I hope America is noticing the difference, because we're making progress. Together, we are changing the tone in the Nation's Capital. And this spirit of respect and cooperation is vital, because, in the end, we will be judged not only by what we say or how we say it, we will be judged by what we're able to do.

America today is a nation with great challenges, but greater resources. An artist using statistics as a brush can paint two very different pictures of our country. One would have warning signs: increasing layoffs, rising energy prices, too many failing schools, persistent poverty, the stubborn vestiges of racism. Another picture would be full of blessings: a balanced budget, big surpluses, a military that is second to none, a country at peace with its neighbors, technology that is revolutionizing the world, and our greatest strength -- concerned citizens who care for our country and care for each other.
likely saved nearly 200 lives. And tonight we welcome and thank flight attendants Hermis Moutardier and Christina Jones. (Applause.)

Once we have funded our national security and our homeland security, the final great priority of my budget is economic security for the American people. (Applause.) To achieve these great national objectives -- to win the war, protect the homeland, and revitalize our economy -- our budget will run a deficit that will be small and short-term, so long as Congress restrains spending and acts in a fiscally responsible manner. (Applause.) We have clear priorities and we must act at home with the same purpose and resolve we have shown overseas: We'll prevail in the war, and we will defeat this recession. (Applause.)

Americans who have lost their jobs need our help and I support extending unemployment benefits and direct assistance for health care coverage. (Applause.) Yet, American workers want more than unemployment checks -- they want a steady paycheck. (Applause.) When America works, America prospers, so my economic security plan can be summed up in one word: jobs. (Applause.)
likely saved nearly 200 lives. And tonight we welcome and thank flight attendants Hermis Moutardier and Christina Jones. (Applause.)

33

Once we have funded our national security and our homeland security, the final great priority of my budget is economic security for the American people. (Applause.) To achieve these great national objectives -- to win the war, protect the homeland, and revitalize our economy -- our budget will run a deficit that will be small and short-term, so long as Congress restrains spending and acts in a fiscally responsible manner. (Applause.) We have clear priorities and we must act at home with the same purpose and resolve we have shown overseas: We'll prevail in the war, and we will defeat this recession. (Applause.)

34

Americans who have lost their jobs need our help and I support extending unemployment benefits and direct assistance for health care coverage. (Applause.) Yet, American workers want more than unemployment checks -- they want a steady paycheck. (Applause.) When America works, America prospers, so my economic security plan can be summed up in one word: jobs. (Applause.)

35
likely saved nearly 200 lives. And tonight we welcome and thank flight attendants Hermis Moutardier and Christina Jones. (Applause.)

Once we have funded our national security and our homeland security, the final great priority of my budget is economic security for the American people. (Applause.) To achieve these great national objectives -- to win the war, protect the homeland, and revitalize our economy -- our budget will run a deficit that will be small and short-term, so long as Congress restrains spending and acts in a fiscally responsible manner. (Applause.) We have clear priorities and we must act at home with the same purpose and resolve we have shown overseas: We'll prevail in the war, and we will defeat this recession. (Applause.)

Americans who have lost their jobs need our help and I support extending unemployment benefits and direct assistance for health care coverage. (Applause.) Yet, American workers want more than unemployment checks -- they want a steady paycheck. (Applause.) When America works, America prospers, so my economic security plan can be summed up in one word: jobs. (Applause.)
likely saved nearly 200 lives. And tonight we welcome and thank flight attendants Hermis Moutardier and Christina Jones. (Applause.)

Once we have funded our national security and our homeland security, the final great priority of my budget is economic security for the American people. (Applause.) To achieve these great national objectives -- to win the war, protect the homeland, and revitalize our economy -- our budget will run a deficit that will be small and short-term, so long as Congress restrains spending and acts in a fiscally responsible manner. (Applause.) We have clear priorities and we must act at home with the same purpose and resolve we have shown overseas: We'll prevail in the war, and we will defeat this recession. (Applause)

Americans who have lost their jobs need our help and I support extending unemployment benefits and direct assistance for health care coverage. (Applause.) Yet, American workers want more than unemployment checks -- they want a steady paycheck. (Applause.) When America works, America prospers, so my economic security plan can be summed up in one word: jobs. (Applause.)
likely saved nearly 200 lives. And tonight we welcome and thank flight attendants Hermis Moutardier and Christina Jones. (Applause.)

Once we have funded our national security and our homeland security, the final great priority of my budget is economic security for the American people. (Applause.) To achieve these great national objectives -- to win the war, protect the homeland, and revitalize our economy -- our budget will run a deficit that will be small and short-term, so long as Congress restrains spending and acts in a fiscally responsible manner. (Applause.) We have clear priorities and we must act at home with the same purpose and resolve we have shown overseas: We'll prevail in the war, and we will defeat this recession. (Applause.)

Americans who have lost their jobs need our help and I support extending unemployment benefits and direct assistance for health care coverage. (Applause.) Yet, American workers want more than unemployment checks -- they want a steady paycheck. (Applause.) When America works, America prospers, so my economic security plan can be summed up in one word: jobs. (Applause.)
we will work with our allies and friends to be a force for good and a champion of freedom. We will work for free markets, free trade and freedom from oppression. Nations making progress toward freedom will find America is their friend. We will promote our values. We will promote the peace. And we need a strong military to keep the peace.

But our military was shaped to confront the challenges of the past. So I've asked the Secretary of Defense to review America's Armed Forces and prepare to transform them to meet emerging threats. My budget makes a down payment on the research and development that will be required. Yet, in our broader transformation effort, we must put strategy first, then spending. Our defense vision will drive our defense budget, not the other way around. (Applause.)

Our nation also needs a clear strategy to confront the threats of the 21st century -- threats that are more widespread and less certain. They range from terrorists who threaten with bombs to tyrants in rogue nations intent upon developing weapons of mass destruction. To protect our own people, our allies and friends, we must develop and we must deploy effective missile defenses. (Applause.)
champion of freedom. We will work for free markets, free trade and freedom from oppression. Nations making progress toward freedom will find America is their friend. We will promote our values. We will promote the peace. And we need a strong military to keep the peace.

But our military was shaped to confront the challenges of the past. So I’ve asked the Secretary of Defense to review America’s Armed Forces and prepare to transform them to meet emerging threats. My budget makes a down payment on the research and development that will be required. Yet, in our broader transformation effort, we must put strategy first, then spending. Our defense vision will drive our defense budget, not the other way around. (Applause.)

Our nation also needs a clear strategy to confront the threats of the 21st century -- threats that are more widespread and less certain. They range from terrorists who threaten with bombs to tyrants in rogue nations intent upon developing weapons of mass destruction. To protect our own people, our allies and friends, we must develop and we must deploy effective missile defenses. (Applause.)
champion of freedom. We will work for free markets, free trade and freedom from oppression. Nations making progress toward freedom will find America is their friend. We will promote our values. We will promote the peace. And we need a strong military to keep the peace.

But our military was shaped to confront the challenges of the past. So I’ve asked the Secretary of Defense to review America’s Armed Forces and prepare to transform them to meet emerging threats. My budget makes a down payment on the research and development that will be required. Yet, in our broader transformation effort, we must put strategy first, then spending. Our defense vision will drive our defense budget, not the other way around. (Applause.)

Our nation also needs a clear strategy to confront the threats of the 21st century — threats that are more widespread and less certain. They range from terrorists who threaten with bombs to tyrants in rogue nations intent upon developing weapons of mass destruction. To protect our own people, our allies and friends, we must develop and we must deploy effective missile defenses. (Applause.)
Trends in distributions

• Given a set of distributions

How to select 'interesting' ones?
Trends in distributions

• Define metrics on distributions and rank features accordingly

  – increase/decrease \(\rightarrow\) topics evolution
Trends in distributions

- Define metrics on distributions and rank features accordingly

- spikes/sinks → specific events

1 spike

1σ
when it turns away from its nuclear ambitions.

Our Nation and the world must learn the lessons of the Korean peninsula, and not allow an even greater threat to rise up in Iraq. A brutal dictator, with a history of reckless aggression...with ties to terrorism...with great potential wealth...will not be permitted to dominate a vital region and threaten the United States.

Twelve years ago, Saddam Hussein faced the prospect of being the last casualty in a war he had started and lost. To spare himself, he agreed to disarm of all weapons of mass destruction. For the next 12 years, he systematically violated that agreement. He pursued chemical, biological, and nuclear weapons even while inspectors were in his country. Nothing to date has restrained him from his pursuit of these weapons - not economic sanctions, not isolation from the civilized world, not even cruise missile strikes on his military facilities. Almost three months ago, the United Nations Security Council gave Saddam Hussein his final chance to disarm. He has shown instead his utter contempt for the United Nations, and for the opinion of the world.
ambitions.

52

Our Nation and the world must learn the lessons of the Korean peninsula, and not allow an even greater threat to rise up in Iraq. A brutal dictator, with a history of reckless aggression...with ties to terrorism...with great potential wealth...will not be permitted to dominate a vital region and threaten the United States.

53

Twelve years ago, Saddam Hussein faced the prospect of being the last casualty in a war he had started and lost. To spare himself, he agreed to disarm of all weapons of mass destruction. For the next 12 years, he systematically violated that agreement. He pursued chemical, biological, and nuclear weapons even while inspectors were in his country. Nothing to date has restrained him from his pursuit of these weapons - not economic sanctions, not isolation from the civilized world, not even cruise missile strikes on his military facilities. Almost three months ago, the United Nations Security Council gave Saddam Hussein his final chance to disarm. He has shown instead his utter contempt for the United Nations, and for the opinion of
neigroors, technology that is revolutionizing the world, and our greatest strength -- concerned citizens who care for our country and care for each other.

Neither picture is complete in and of itself. And tonight I challenge and invite Congress to work with me to use the resources of one picture to repaint the other; to direct the advantages of our time to solve the problems of our people. Some of these resources will come from government. Some, but not all.

Year after year in Washington, budget debates seem to come down to an old, tired argument: on one side, those who want more government, regardless of the cost; on the other, those who want less government, regardless of the need. We should leave those arguments to the last century, and chart a different course. (Applause.)

Government has a role, and an important role. Yet, too much government crowds out initiative and hard work, private charity and the private economy. Our new governing vision says government should be active, but limited; engaged, but not overbearing. And my budget is based on that philosophy.
Long patterns
The United Nations concluded in 1999 that Saddam Hussein had biological weapons materials sufficient to produce over 25,000 liters of anthrax - enough doses to kill several million people. He has not accounted for that material. He has given no evidence that he has destroyed it.

The United Nations concluded that Saddam Hussein had materials sufficient to produce more than 38,000 liters of botulinum toxin - enough to subject millions of people to death by respiratory failure. He has not accounted for that material. He has given no evidence that he has destroyed it.

Our intelligence officials estimate that Saddam Hussein had the materials to produce as much as 500 tons of sarin, mustard, and VX nerve agent. In such quantities, these chemical agents also could kill untold thousands. He has not accounted for these materials. He has given no evidence that he has destroyed them.

U.S. intelligence indicates that Saddam Hussein had upwards of 30,000 munitions capable of delivering chemical agents. Inspectors recently turned up 16 of them, despite Iraq's recent declaration denying their existence. Saddam Hussein has not accounted for the remaining 29,984 of these prohibited munitions. He has given no evidence that he has destroyed them.

From three Iraqi defectors we know that Iraq, in the late 1990s, had several mobile biological weapons labs. These are designed to produce germ warfare
• n-grams (n consecutive words)

exact 10-gram

<table>
<thead>
<tr>
<th>Doc 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>My fellow citizens, I will improve security in our country</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Doc 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>My fellow citizens, I will improve security in our country</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Doc 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>My fellow citizens, I will improve security in our country</td>
</tr>
</tbody>
</table>

Catches exact matches
• Frequent sets of small n-grams

Pattern of four 3-grams

fellow citizens I / citizens I will / I will improve / in our country

Doc 1
My fellow citizens, I will improve social welfare in our country

Doc 2
My fellow citizens, I will improve medical aid in our country

Doc 3
Dear fellow citizens, I will improve security in our country

Catches “fuzzier” matches
The United Nations concluded in 1999 that Saddam Hussein had biological weapons materials sufficient to produce over 25,000 liters of anthrax - enough doses to kill several million people. He has not accounted for that material. He has given no evidence that he has destroyed it.

The United Nations concluded that Saddam Hussein had materials sufficient to produce more than 38,000 liters of botulinum toxin - enough to subject millions of people to death by respiratory failure. He has not accounted for that material. He has given no evidence that he has destroyed it.

Our intelligence officials estimate that Saddam Hussein had the materials to produce as much as 500 tons of sarin, mustard, and VX nerve agent. In such quantities, these chemical agents also could kill untold thousands. He has not accounted for these materials. He has given no evidence that he has destroyed them.

U.S. intelligence indicates that Saddam Hussein had upwards of 30,000 munitions capable of delivering chemical agents. Inspectors recently turned up 16 of them, despite Iraq's recent declaration denying their existence. Saddam Hussein has not accounted for the remaining 29,984 of these prohibited munitions. He has given no evidence that he has destroyed them.

From three Iraqi defectors we know that Iraq, in the late 1990s, had several mobile biological weapons labs. These are designed to produce nerve warfare...
Evaluation strategy

• Ongoing longitudinal case study
  – Tanya Clement and « The Making of Americans »

• Pilot user study with 8 users
  – 3 tasks then free exploration (30 min)
  – think aloud protocol - gather insights about text and + functionality used
  – User rated insights found for interest (1-5 scale).
was working to make some women understand something and many laughed at her and always she was full of desiring and always she was never understanding in desiring. When there was the end of her living with Redfern her brother Alfred was just coming to his marrying Julia Dehning. Martha was then travelling and studying and then she came back to be with her father and her mother was weakening then and later she was dead and Mr.

1808

Hersland lost his great fortune and Martha then took care of him. There will be now a little more description of her and then of her with him. There will be a little more description of her written in the history of the ending of the living in her father, in the history of the later living of her brother Alfred Hersland, in the history of her brother David Hersland. More description of her will be part of the history of the ending of the existing of the Hersland family. There will be very much history of this ending of all of them of the Hersland family written later.

1809

There will be now a little more description written of her and of her living with her father when she came back to the family living back out of her trouble after the ending of the living in Phillip Redfern.
At first filled him with astonishment and then with delight. He could not feel himself a part of it, he could not love the sense of danger in the presence and companionship of women, his instincts bade him be on guard but his ideal he felt to be here realised.

Among the many vigorous young women in the place there was Martha Hersland. She was a blond good-looking young woman full of moral purpose and educational desires. She had an eager earnest intelligence, fixed convictions and principles by then, and restless energy. She and Redfern were students in the same studies in the same class and soon singled themselves out from the crowd, it was all new, strange and dangerous for the south-western man and all perfectly simple and matter of course for the western girl. They had long talks on the meanings of things, he discoursing of his life and aims, she listening, understanding and sympathising. This intercourse steadily grew more constant and familiar. Redfern's instincts were dangerous was always there as a conviction in him, his ideals simple and pure was almost always real inside him, slowly he realised in this constant companion the existence of instincts as simple and pure as his ideals.
was working to make some women understand something and many laughed at her and always she was full of desiring and always she was never understanding in desiring. When there was the end of her living with Redfern her brother Alfred was just coming to his marrying Julia Dehnring. Martha was then travelling and studying and then she came back to be with her father and her mother was weakening then and later she was dead and Mr.

Hersland lost his great fortune and Martha then took care of him. There

description written of her and of her living with her father when she came back to the family living back out of her trouble after the ending of the living in Phillip Redfern.
Collection Overview: 'The Making of Americans' (5 doc per line - 311 hits)

Sections Overview

Legend:
- Red: completely
- Green: he was not
- Orange: was not completely
- Blue: was almost completely

Document View

was almost completely clearly feeling being one being living. He was almost completely clearly feeling in being one being living. He was one almost completely clearly feeling about needing being one being living, about not needing being one being living, about knowing being one being living, about being certain of being one being living, about not being certain of being one being living. He was almost completely clearly feeling something. He was almost completely clearly feeling everything of any such thing.

2948

He was almost completely clearly feeling that someone was different from any other one. He was almost completely clearly feeling such a thing. He was certainly mentioning this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was mentioning being certain that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was almost completely clearly feeling something about this thing about
was almost completely clearly feeling being one being living. He was almost completely clearly feeling in being one being living. He was one almost completely clearly feeling about needing being one being living, about not needing being one being living, about knowing being one being living, about being certain of being one being living, about not being certain of being one being living. He was almost completely clearly feeling something, he was almost completely clearly feeling everything of any such thing.

2948

He was almost completely clearly feeling that someone was different from any other one. He was almost completely clearly feeling such a thing. He was certainly mentioning this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was almost completely clearly feeling something about this thing about
Collections tested

• Making of the Americans
  – 8 sections, ~ 3000 paragraphs

• State of the union
  – 7 years of speeches

• HCIL Tech Reports
  – 20 years of lists of abstracts

• Other book: Gamer Theory
Implementation

• Open Lazlo for client user interface
  – Zero install access
• T2K (Text-to-Knowledge) for text mining
  – Thanks to Loretta Auvil’s team at NSCA
  – Pre-processing
• Ruby and MySQL for the backend part
• Open Source
Questions we addressed

• What text features are highly repeated in the text?
  – Frequent words
  – Frequent patterns of n-grams

• How do they change over time?
  – Locate features in text (overview + reading)
  – Comparisons
  – Study the shape of change
  – Find features that exhibit specific shapes (e.g. spike)
Next

• What text features are highly repeated in the text?
  – Frequent words
  – Frequent patterns of n-grams

• How do they change over time?
  – Locate features in text (overview + reading)
  – Comparisons
  – Study the shape of change
  – Find features that exhibit specific shapes (e.g. spike)
Thank you

• Text mining results can be made easier to interpret
• Repeating is everywhere
• Technical Report HCIL 2007-08
• Talk to Tanya at the demo

Anthony Don, Catherine Plaisant, Tanya Clement
Elena Zheleva, Machon Gregory, Sureyya Tarka,
Celeste Paul, Ben Shneiderman

and others from the MONK project, in particular Loretta Auvil and Matt Kirschenbaum

Support from the Andrew W. Mellon Foundation
at first filled him with astonishment and then with delight. He could not feel himself a part of it, he could not love the sense of danger in the presence and companionship of women, his instincts bade him be on guard but his ideal he felt to be here realised.

1684

Among the many vigorous young women in the place there was Martha Hersland. She was a blond good-looking young woman full of moral purpose and educational desires. She had an eager earnest intelligence, fixed convictions and principles by then, and restless energy. She and Redfern were students in the same studies in the same class and soon singled themselves out from the crowd, it was all new, strange and dangerous for the south-western man and all perfectly simple and matter of course for the western girl. They had long talks on the meanings of things, he discoursing of his life and aims, she listening, understanding and sympathising. This intercourse steadily grew more constant and familiar. Redfern’s instincts were dangerous was always there as a conviction in him, his ideals simple and pure was almost always real inside him, slowly he realised in this constant companion the existence of instincts as simple and pure as his ideals.

1685
at first filled him with astonishment and then with delight. He could not feel himself a part of it, he could not love the sense of danger in the presence and companionship of women, his instincts bade him be on guard but his ideal he felt to be here realised.

Among the many vigorous young women in the place there was Martha Hersland. She was a blond good-looking young woman full of moral purpose and educational desires. She had an eager earnest intelligence, fixed convictions and principles by then, and restless energy. She and Redfern were students in the same studies in the same class and soon singled themselves out from the crowd, it was all new, strange and dangerous for the south-western man and all perfectly simple and matter of course for the western girl. They had long talks on the meanings of things, he discoursing of his life and aims, she listening, understanding and sympathising. This intercourse steadily grew more constant and familiar. Redfern's instincts were dangerous was always there as a conviction in him, his ideals simple and pure was almost always real inside him, slowly he realised in this constant companion the existence of instincts as simple and pure as his ideals.
was working to make some women understand something and many laughed at her and always she was full of desiring and always she was never understanding in desiring.

When there was the end of her living with Redfern her brother Alfred was just coming to his marrying Julia Dehnring. Martha was then travelling and studying and then she came back to be with her father and her mother was weakening then and later she was dead and Mr.

1808

Hersland lost his great fortune and Martha then took care of him. There will be now a little more description of her and then of her with him. There will be a little more description of her written in the history of the ending of the living in her father, in the history of the later living of her brother Alfred Hersland, in the history of her brother David Hersland. More description of her will be part of the history of the ending of the existing of the Hersland family. There will be very much history of this ending of all of them of the Hersland family written later.

1809

There will be now a little more description written of her and of her living with her father when she came back to the family living back out of her trouble after the ending of the living in Phillip Redfern.

1810
was working to make some women understand something and many laughed at her and always she was full of desiring and always she was never understanding in desiring. When there was the end of her living with Redfern her brother Alfred was just coming to his marrying Julia Dehning. Martha was then travelling and studying and then she came back to be with her father and her mother was weakening then and later she was dead and Mr.

1808

Hersland lost his great fortune and Martha then took care of him. There will be now a little more description of her and then of her with him. There will be a little more description of her written in the history of the ending of the living in her father's, in the history of the later living of her brother Alfred Hersland, in the history of her brother David Hersland. More description of her will be part of the history of the ending of the existing of the Hersland family. There will be very much history of ending of all of them of the Hersland family written later.

1809

There will be now a little more description written of her and of her living with her father when she came back to the family living back out of her trouble after the ending of the living in Phillip Redfern.
was working to make some women understand something and many laughed at her and always she was full of desiring and always she was never understanding in desiring. When there was the end of her living with Redfern her brother Alfred was just coming to his marrying Julia Dehning. Martha was then travelling and studying and then she came back to be with her father and her mother was weakening then and later she was dead and Mr.

Hersland lost his great fortune and Martha then took care of him. There will be now a little more description of her and then of her with him. There will be a little more description of her written in the history of the ending of the living in her father, in the history of the later living of her brother Alfred Hersland, in the history of her brother David Hersland. More description of her will be part of the history of the ending of the existing of the Hersland family. There will be very much history of this ending of all of them of the Hersland family written later.

There will be now a little more description written of her and of her living with her father when she came back to the family living back out of her trouble after the ending of the living in Phillip Redfern.
Text mining

Features: words, n-grams, patterns

Distribution of Features over time

Profiling of distribution

Ranked features

User Interface
He was almost **completely** clearly feeling that someone was different from any other one. He was almost **completely** clearly feeling such a thing. He was certainly mentioning this thing that someone was different from any other one. He was not **completely** certain of this thing that someone was different from any other one. He was mentioning being certain that someone was different from any other one. He was not **completely** clearly feeling something about this thing about someone being different from any other one. He went on being living until the ending of the beginning of middle living.
was almost completely clearly feeling being one being living. He was almost completely clearly feeling in being one being living. He was one almost completely clearly feeling about needing being one being living, about not needing being one being living, about knowing being one being living, about being certain of being one being living, about not being certain of being one being living. He was almost completely clearly feeling something. He was almost completely clearly feeling everything of any such thing.

2948

He was almost completely clearly feeling that someone was different from any other one. He was almost completely clearly feeling such a thing. He was certainly mentioning this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was not completely certain of this thing that someone was different from any other one. He was almost completely clearly feeling something about this thing about
Once an angry man dragged his father along the ground through his own orchard. "Stop!" cried the groaning old man at last. "Stop! I did not drag my father beyond this tree."

It is hard living down the tempers we are born with. We all begin well, for in our youth there is nothing we are more intolerant of than our own sins writ large in others and we fight them fiercely in ourselves; but we grow old and see that these our sins are of all sins the really harmless ones to own, nay that they give a charm to any character, and so our struggle with them dies away.

It has always seemed to me a rare privilege, this, of being an American, a real American, one whose tradition it has taken scarcely sixty years to create. We need only realize our parents, remember our grandparents and know ourselves and our history is complete.

The old people in a new world, the new people made out of the old, that is the story that I mean to tell, for that is what really is and what I really know.

Some of the fathers we must realize so that we can tell our story really, were little boys then, and they came
DEMO WITH ranked distributions
Example of insights found using the spike ranking
Sample of data mining results

Everyone then is an individual being. Everyone then is like many others always living thinking of everyone, this is not a description of all of them. There must then be a whole history must then now be a description of all repeating. Now I will tell all the meaning to me in repeating repeating.

Everyone is one inside them, everyone reminds someone of some other one who is. Everyone has it to say of each one he is like such a one I see it in him, everyone has it to say of each else I can tell by remembering. So it goes on always in living, everyone is always remembering some the one at whom they are then looking. So they go on repeating, everyone is themselves inside the resembling to others, and that is always interesting. There are many ways of making kinds of men, making kinds of them there is a different system of finding them resembling. Sometimes there will be of seeing kinds of men and women. Sometimes there will be then a complete history of each or repeating the whole of them and so sometime someone who sees them will have a complete history someone will know all the ways there are for people to be resembling, someone sometime then will of everyone.

Soon now there will be a history of the way repeating comes out of them comes out of them. They are young, when they are children, they have their own system of being resembling, the men and women in beginning, the being young in them, the being children.

There is then now and here the loving repetition, this is then, now and here, a description of repetition and then there will be a description of all the kinds of ways there can be seen to be kinds of ways there will be realised the complete history of everyone, the fundamental character of everyone, their mixtures in them, the strength and weakness of everything they have inside them, the flavor of the being in them, and then you have a whole history then of each one. Everything then they do in living understanding, their living, loving, eating, pleasing, smoking, thinking, scolding, drinking, working, laughing, sleeping, everything in them. There are whole beings then, they are themselves inside them.
Sample of data mining results

5: a description of
4: this is now + is now a
4: there will be
4: of men and + men and women
4: of each one
4: men and women
4: is now a
4: history of each + of each one
3: this is now + a description is now a description of
3: there will be + will be a + a description of
3: there will be of each one
3: there will be a description of
3: there are many
3: one of them history of each one of each one
3: of men and there will be
3: of men and men and women kinds of men
3: of men and men and women a history of
3: of each one a description of
3: now i will
3: men and women there will be
3: is now a description of
3: history of each there will be
3: history of each one of each one a description of
3: history of each a description of
3: be a description a description of
2: will be a of each one
2: this is now there are many is now a now i will.

148 patterns (4 pages) ← 1 page of MoA

<p n="1082">Everyone then is an individual being. Everyone then is like many others always living thinking of everyone, this is now a description of all of them. There must then be a whole history must then now be a description of all repeating. Now I will tell all the meaning to me in repeating repeating.</p>

<p n="1083">Everyone is one inside them, everyone reminds someone of some other one who is. Everyone has it to say of each one he is like such a one I see it in him, everyone has it to say of everyone else I can tell by remembering. So it goes on always in living, everyone is always remembering someone the one at whom they are then looking. So they go on repeating, everyone is themselves inside the resembling to others, and that is always interesting. There are many ways of making kinds of men making kinds of them there is a different system of finding them resembling. Sometime there will be of seeing kinds of men and women. Sometime there will be then a complete history of each or repeating the whole of them and so sometime someone who sees them will have a complete history someone will know all the ways there are for people to be resembling, someone sometime then will of everyone.</p>

<p n="1084">Soon now there will be a history of the way repeating comes out of them comes out. If they are young, when they are children, they have then their own system of being resembling, then the men and women in beginning, the being young in them, the being children.</p>

<p n="1085">There is then now and here the loving repetition, this is then, now and here, a description repetition and then there will be a description of all the kinds of ways there can be seen to be kind there will be realized the complete history of everyone, the fundamental character of everyone, their mixtures in them, the strength and weakness of everything they have inside them, the flavor of them being in them, and then you have a whole history then of each one. Everything then they do in living understanding, their living, loving, eating, pleasing, smoking, thinking, scolding, drinking, working, laughing, sleeping, everything in them. There are whole beings then, they are themselves inside them.
Sample of data mining results

5: a description of (7 occurrences, in 5 paragraphs, also in 41 patterns!)

4: this is now + is now a (4 occurrences in the same paragraphs)
4: of men and + men and women
4: of each one (9 occurrences in 4 paragraphs)
4: men and women (6 occurrences in 4 paragraphs)
4: is now a (6 occurrences in 4 paragraphs)
4: history of each + of each one

3: this is now + now a description + is now a + a description of <<<<<REALLY all occurrences at “This is now a description of” so longer Ngrams are more useful than)

3: this is now + a description of <<<<<BETTER cause are ADJACENT
3: there will be + will be a + a description of
3: there will be of each one
3: there will be a description of
3: there are many
3: one of them history of each each one of of each one a description of
3: of men and there will be
3: of men and men and women kinds of men
3: of men and men and women a history of
3: of each one a description of
3: now i will
3: men and women there will be
Sample of data mining results

5: a description of (7 occurrences, in 5 paragraphs, also in 41 patterns!)

4: this is now + is now a (4 occurrences in the same paragraphs)
<><><> THOSE ALWAYS CO-OCCUR (and always ADJACENT i.e. a 4-gram)
4: there will be (6 occurrences in 4 paragraphs)
4: of men and + men and women
4: of each one (9 occurrences in 4 paragraphs)
4: men and women (6 occurrences in 4 paragraphs)
4: is now a (6 occurrences in 4 paragraphs)
4: history of each + of each one

3: this is now + now a description + is now a + a description of<><><>REALLY all occurrences at “This is now a description of” so longer Ngrams more useful that than.

3: this is now + a description of<><><>BETTER cause are ADJACENT
3: there will be + will be a + a description of
3: there will be of each one
3: there will be a description of
3: there are many
3: one of them history of each each one of of each one a description of
3: of men and there will be
3: of men and men and women kinds of men
3: of men and men and women a history of
3: of each one a description of
3: now i will
3: men and women there will be
Sample of data mining results

5: a description of (7 occurrences, in 5 paragraphs, also in 41 patterns!)

4: this is now + is now a (4 occurrences in the document - THIS ALWAYS CO-OCCUR (and always the same) + those)

4: there will be (6 occurrences in 4 paragraphs)

4: of men and + men and women (9 occurrences in 4 paragraphs)

4: is now a (6 occurrences in 4 paragraphs)

4: history of each + of each one

3: this is now + now a description + is now a + a description of occurrences at “This is now a description of” so long

3: this is now + a description of <<<<<BETTER

3: there will be + will be a + a description of

3: there will be of each one

3: there will be a description of

3: there are many

3: one of them history of each each one of of each one of

3: of men and there will be

3: of men and men and women kinds of men

3: of men and men and women a history of

3: of each one a description of

3: now i will

3: men and women there will be

<p>Everyone then is an individual being. Everyone then is like many others always living, thinking of everyone, this is now a description of all of them. There must then be a whole history, must then be a description of all repeating. Now I will tell all the meaning to me in repeating repeating.</p>

<p>Everyone is one inside them, everyone reminds someone of some other one who is. Everyone has it to say of each one he is like such a one I see it in him, everyone has it to say of a else I can’t tell by remembering. So it goes on always in living, everyone is always remembering a the one at whom they are then looking. So they go on repeating, everyone is themselves inside the resembling to others, and that is always interesting. There are many ways of making kinds of men making kinds of them there is a different kind of finding them resembling. Sometime there will be of seeing kinds of men and women. Sometime there will be then a complete history of each of repeating the whole of them and so sometime someone who sees them will have a complete history of someone will know all the ways there are for people to be resembling, someone sometime will of</p>

<p>Soon now there will be a history of the way repeating comes out of them come out of they are young, when they are children, they have then their own system of being resembling, this the men and women in beginning, the being young in them, the being children.</p>

<p>There is then now and here the living repetition, this is then, now and here, a deep repetition and then there will be a description of all the kinds of ways there can be seen to be kind such will be realised the complete history of everyone, the fundamental character of everyone, the mixture in them, the strength and weakness of everything they have inside them, the flavor of them, being in them, and then you have a whole history there of each one. Everything then they do in life understanding, their living, loving, eating, pleasing, smoking, thinking, scolding, drinking, works, laughing, sleeping, everything in them. There are whole beings then, they are themselves inside the whole of them makes a history of each one of them.</p>

<p>Always from the beginning there was to me all living as repeating. This is now a description was saying listening to repeating is often irritating, always repeating is all of living, everything in more and more listening to repeating gives to me completely understanding.</p>