Name This! Automating Metadata Extraction through a Named Entity Recognition Tool

Jean Godby - OCLC

Judith Klavans - University of Maryland

Dan Roth - University of Illinois at Urbana-Champaign

Patricia Hswe - University of Illinois at Urbana-Champaign

Presentation Outline

- The problem of named entity recognition

- Collaboration among three institutions
  - University of Illinois at Urbana-Champaign
  - University of Maryland
  - OCLC

Three problems

For each name in a stream of text:

- **Recognize it**
  - Distinguish names from non-names
  - Assign it to a broadly recognized category
    - Personal, corporate, geographic

- **Cluster it**
  - Associate variants of the same name

- **Assign an identity**
  - … or the name's real-world referent
  - Select the canonical form of a name

An example

Kennedy

Document 1: *The Justice Department has officially ended its inquiry into the assassinations of* John F. Kennedy *and Martin Luther King Jr., finding `no persuasive evidence" to support conspiracy theories, according to department documents. The House Assassinations Committee concluded in 1978 that* Kennedy *was `probably" assassinated as the result of a conspiracy involving a second gunman, a finding that broke from the* **Warren Commission**'s *belief that Lee Harvey Oswald acted alone in* **Dallas** *on Nov. 22, 1963.*

Document 2: *In 1953, Massachusetts* Sen. John F. Kennedy *married Jacqueline Lee Bouvier in Newport, R.I. In 1960, Democratic presidential candidate* John F. Kennedy *confronted the issue of his Roman Catholic faith by telling a Protestant group in Houston, ``I do not speak for my church on public matters, and the church does not speak for me. "* JFK *was known for his liberal approaches to religion and civil rights.*

Document 3: David Kennedy *was born in Leicester, England in 1959.* ... Kennedy *co-edited The New Poetry (Bloodaxe Books 1993), and is the author of New Relations: The Refashioning Of British Poetry 1980-1994 (Seren 1996).*

- **Recognize it**
  - *John F. Kennedy*
  - [PERS John F. Kennedy]
  - [PERS David Kennedy]
- **Cluster it**
  - *John F. Kennedy, Kennedy*
  - *David Kennedy, Kennedy*
  - John F. Kennedy != David Kennedy
- **Assign an identity**
  - John F. Kennedy, President and Senator
  - David Kennedy, Poet and Author

EMP: *Extracting Metadata for Preservation*

- NDIIPP National Digital Information Infrastructure and Preservation Program
- Funded by the Library of Congress
- Grant to University of Illinois at Urbana-Champaign.
- University of Maryland and OCLC are subcontractors.

EMP project goals

– Lower the barrier of access to high-end Named-Entity Recognition tools

– Create tools for open use

– Demonstrate use in digital library applications

Project deliverables

1. Use named entity recognition (NER) toolkit developed at the University of Illinois
2. Address the three problems with named entities.
    –   automatically extract proper names from text
    –   normalize them
    –   develop algorithms to assign identities
3. Evaluate with human subjects.
4. Create a Web service interface for the NER tool.
5. Demonstrate the NER toolkit in library applications.
6. Deliver the toolkit as an open source software package.

Three Institution Collaboration

1. Named Entity Recognition (NER)

   - University of Illinois at Urbana-Champaign

2. Computational Linguistics for Metadata Building (CLiMB)

   University of Maryland

3. Research projects on names management

   - OCLC

The Named Entity Recognizer

- Developed by Dan Roth at UIUC (Cognitive Computation Group, Department of Computer Science)

- Implements a machine learning solution to named entity recognition

- Demo available at:
  http://l2r.cs.uiuc.edu/~cogcomp/LbjNer.php

Cognitive Computation Group

Named Entity Tagger Output

The input sentences are:

o George Washington (February 22, 1732[1][2][3] – December 14, 1799) was the first President of the United States, (1789–1797),[4] and led the Continental Army to victory over the Kingdom of Great Britain in the American Revolutionary War (1775–1783). Washington was chosen to be the commander-in-chief of the American revolutionary forces in 1775. The following year, he forced the British out of Boston, lost New York City, and crossed the Delaware River in New Jersey and defeated the surprised enemy units later that year. As a result of his strategy, Revolutionary forces captured the two main British combat armies at Saratoga and Yorktown. Negotiating with Congress, the colonial states, and French allies, he held together a tenuous army and a fragile nation amid the threats of disintegration and failure. Following the end of the war in 1783, Washington retired to his plantation on Mount Vernon, prompting even King George III to name him "the greatest man in the world."[5][need quote

The tagged output is:

o [PER George Washington] (February 22 , 1732 (1) (2) (3) Â– December 14 , 1799) was the first President of the [LOC United States] , (1789 Â– 1797) , (4) and led the [ORG Continental Army] to victory over the [ORG Kingdom] of Great [LOC Britain] in the [MISC American Revolutionary War] (1775 Â– 1783) . [LOC Washington] was chosen to be the commander-in-chief of the [MISC American] revolutionary forces in 1775 . The following year , he forced the [MISC British] out of [LOC Boston] , lost [LOC New York City] , and crossed the [LOC Delaware River] in [LOC New Jersey] and defeated the surprised enemy units later that year . As a result of his strategy , Revolutionary forces captured the two main [MISC British] combat armies at [LOC Saratoga] and [ORG Yorktown] . Negotiating with [ORG Congress] , the colonial states , and [MISC French] allies , he held together a tenuous army and a fragile nation amid the threats of disintegration and failure . Following the end of the war in 1783 , [LOC Washington] retired to his plantation on [PER Mount Vernon] , prompting even King [PER George III] to name him " the greatest man in the world. " (5) (need quote

Key:

Under development for EMP

- New research on identity resolution

- Increase usability
  - Web service interface
  - Workflows for training
  - Incorporating data into demos

- Performance evaluation and tuning
  - Comparisons with human taggers
  - Comparisons with other algorithms
  - Evaluation of scoring methodologies

Three Institution Collaboration

1. Named Entity Recognition (NER)
   - University of Illinois at Urbana-Champaign

2. Computational Linguistics for Metadata Building (CLiMB)

   University of Maryland

3. Research projects on names management
   - OCLC

Home   **Videos**   **Shows**   **Channels**   **Community**

john kennedy funeral bobby   **Search**   **Upload**

Clip of the John F. Kennedy funeral

0:10 / 1:22   HQ

★★★★★ 47 ratings

55,468 views

♥ **Favorite**   ➜ **Share**   ✚ **Playlists**   ⚑ **Flag**

MySpace   Facebook   Twitter   (more share options)

▶ **Statistics & Data**

▶ **Video Responses** (0)   Sign in to post a Video Response

▼ **Text Comments** (29) Options   Sign in to post a Comment

sonic1285 (2 months ago)   Reply   +1 👎 👍
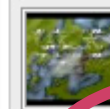
oldnewsclipster
May 29, 2007

(less info)

Key moments during the John F. Kennedy funeral.
Shows Bobby Kennedy, Marines folding the United
States Flag, and the eternal flame.

Category: News & Politics

Tags: Jacqueline Kennedy Assassination
President national tragedy victory for
stupidity

URL   http://www.youtube.com/watch?v=LOIzTLq

Embed   <object width="425" height="344"><param ⚙

▶   **More From: oldnewsclipster**

▼   **Related Videos**

JFK Funeral
102,864 views
GJJdude

R.F.K.
44,386 views
eggman54
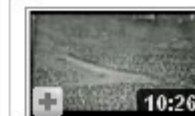
**The Funeral of John F.
Kennedy Part 1**
19,680 views
armysaber

Funeral Service at the Grave
1,527,095 views
ReaderWilliamM11

**1945 President Roosevelt
Funeral**

Sign Up | QuickList (0) | Help | Sign In

Home | Videos | Shows | Channels | Community

john kennedy funeral

Search

Upload

The Funeral of John F. Kennedy in colour

HelmerReenberg
December 17, 2008
(less info)

Subscribe

On Sunday afternoon, November 24th 1963, about 300,000 people watched a horse-drawn caisson, which had borne the body of Franklin D. Roosevelt and the Unknown Soldier, carry Kennedy's flag-covered mahogany casket down the White House drive, past parallel rows of soldiers bearing the flags of the 50 states of the Union, then along Pennsylvania Avenue to the Capitol Rotunda to lie in state. The only sounds on Pennsylvania Avenue as the cortège made its way to the Capitol were the sounds of the muffled drums and the clacking of horses' hooves.

Monday, November 25th 1963, the casket was borne again by caisson on the final leg to Arlington National Cemetery for burial. Moments after the casket was carried down the front steps of the cathedral, Jacqueline Kennedy whispered to her son, after which he saluted his father's coffin, a gesture captured by the cameras and long remembered. The children were deemed to be too young to attend the final burial service, so this was the point where the children said goodbye to their father.

Category: News & Politics

Tags: JFK Assassination Jacqueline Kennedy Caroline Robert John jr.

URL http://www.youtube.com/watch?v=AiqrUMv

Embed <object width="425" height="344"><param

08:40 / 10:49

Looking for Bobby Kennedy

♥ Favorite

MySpace

▶ Statistics &

▶ Video Responses (0)

Sign in to post a Video Response

▼ Text Comments (20) Options

Sign in to post a Comment

Mobilecheese (15 hours ago)

Ok, I'll take your word for it. But if they can alter JFK's wounds during the

Reply 0 👎 👍

▶ More From: HelmerReenberg

▼ Related Videos

The Funeral of John F.

CLiMB
Computational Linguistics for Metadata Building

- Use text associated with images
- Assist cataloger in identifying high quality subject descriptors
- Uses a range of computational analysis tools
- Named Entities are linked to existing resources

The Problem:

Insufficient Subject Access to Images

Main Entry: Category or Artist

Egypt (New Kingdom)

Title or Description

Queen Nefertiti: (front view), ca. 1350 B.C. [Dynasty XVIII]

[Location] • AICT Photo ID #

Aegiptisches Museum, Berlin
#EN010

Text References

Adams 3: pl. 5.22 [alt.]
Adams 4: pl. 5.22 [alt.]
Adams AAT: pl. 4.42 [alt.]
Adams AAT2: pl. 3.38 [alt.]
Gardner 11: pl. 3-33
Gardner 12: pl. 3-33
Hartt 4: pl. 3-42
H&F 4: pl. 3.16 [alt.]
H&F 5: pl. 3.16 [alt.]
Janson 5: pl. 82 [alt.]
Janson 5 R: pl. 79
Janson 6: pl. 2-28 [alt.]
Stokstad: pl. 3-38
Stokstad R: pl. 3-38
Stokstad 2: pl. 3-38

To download image:

Mac: hold down option and click photo size.

PC: hold down command and click desired image size.

x250 pixels
or
x512 pixels

start · 5 Netscape · Microsoft PowerPoint ... · untitled - Paint

Current Process Flow
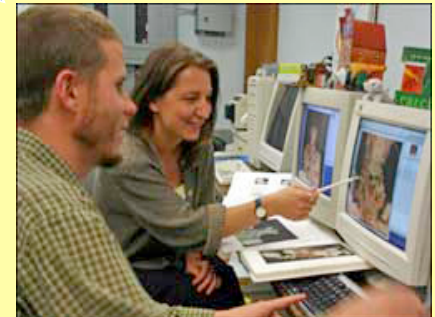
Nefertiti

Gardner (v. 11, pl. 3-33)

The famous painted limestone bust of Akhenaton's queen, Nefertiti (fig. 3-33), exhibits a similar expression of entranced musing and an almost mannered sensitivity and delicacy of curving contour. The piece was found in the workshop of the queen's official sculptor, Thutmose, and is a deliberately unfinished model very likely by the master's own hand. The left eye socket still lacks the inlaid eyeball, making the portrait a kind of before-and-after demonstration piece. With this elegant bust, Thutmose may have been alluding to a heavy flower on its slender stalk by exaggerating the weight of the crowned head and the length of the almost serpentine neck…

Nefertiti

The famous painted limestone bust of Akhenaton's queen, Nefertiti (fig. 3-33), exhibits a similar expression of entranced musing and an almost mannered sensitivity and delicacy of curving contour. The piece was found in the workshop of the queen's official sculptor, Thutmose, and is a deliberately unfinished model very likely by the master's own hand. The left eye socket still lacks the inlaid eyeball, making the portrait a kind of before-and-after demonstration piece. With this elegant bust, Thutmose may have been alluding to a heavy flower on its slender stalk by exaggerating the weight of the crowned head and the length of the almost serpentine neck…

Image Catalogers

Catalog Record
with Enhanced Subject Metadata

Main Entry: Category or Artist

Egypt (New Kingdom)

Title or Description

Queen Nefertiti: (front view),
ca. 1350 B.C. [Dynasty XVIII]

[Location] • AICT Photo ID #

Aegiptisches Museum, Berlin
#EN010

CLiMB Terms

limestone [carbonate rock], aat id: 300011286
busts [figures], aat id: 300047457
Akhenaton
queens [monarchs], aat id: 300025483

more CLiMB terms

Text References

Adams 3: pl. 5.22 [alt.]
Adams 4: pl. 5.22 [alt.]
Adams AAT: pl. 4.42 [alt.]
Adams AAT2: pl. 3.38 [alt.]
Gardner 11: pl. 3-33
Gardner 12: pl. 3-33

Image Searchers
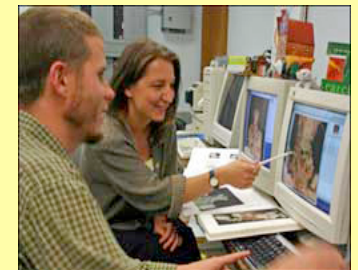
CLiMB Approach

Techniques exceed simple keyword indexing by:

1. applying advanced semantic categorization to text segments
2. identifying coherent phrases
3. associating terms with a thesaurus, and
4. applying disambiguation algorithms to these terms.

The CLiMB Interface

Under development for EMP

- Compare performance of UIUC name extractor tool and CLiMB's name extractor

- Evaluation and usability studies

- Research on scoring evaluation

Three Institution Collaboration

1. Named Entity Recognition (NER)
   - University of Illinois –Urbana Champagne

2. Computational Linguistics for Metadata Building (CLiMB)

   University of Maryland

3. Research projects on names management
   - OCLC

OCLC

- Our needs
  - Subject access to the published record
  - Improved authority files
  - Links to unstructured and un-curated text
- Our resources
  - WorldCat Identities
  - Linked authority files
  - Growing collections of unstructured text

WorldCat® Identities [Beta]

Matches for *'kennedy, john f.'*

1. Kennedy, John F. published: 1940-2008 (Children of presidents--United States)

2. Kennedy, John F. (John Fitzgerald) 1917-1963 (Presidents--United States)

3. Kennedy, John F. (John Fitzgerald) 1960-1999 (Children of presidents--United States)

4. John F. Kennedy Center for the Performing Arts (U.S.) published: 1959-2008

5. John F. Kennedy Space Center published: 1957-2008

6. John F. Kennedy School of Government published: 1965-2008

7. John F. Kennedy Library published: 1954-2008

8. Kennedy, John F. 1942-

Results Key

- A Personal Identity
- A Personal Identity from a controlled vocabulary
- A Corporate Identity
- A Corporate Identity from a controlled vocabulary

Text Size indicates relative popularity of an Identity

(Subject) The information in parentheses indicate the Subject Heading associated with that Identity

WorldCat® Identities

Jump To: Overview | Publication Tim...

Kennedy, John

Overview

Works: 7,205 works i

Genres: Speeches, a

Subject Headings: Presidents--U

Roles: Performer, S

Classifications: e176, 973.09

Publication Timeline

1930

Key
- Publications about John
- Publications by John F
- Publications by John F

- **Onassis, Jacqueline Kennedy**
- **King, Martin Luther Jr 1929-1968** Performer

Useful Links

- Library of Congress Authority File (English)
- Deutsche Nationalbibliothek Authority File (German)
- Wikipedia John F. Kennedy

Associated Subjects

1900 - 1999   1939 - 1945   1961 - 1963   1961 - 1969   1962   African Americans--Civil rights   Archives   Assassination   Assassins   Biography   Case studies   Conspiracies   Courage   Cuba   Cuban Missile Crisis (1962)   Family   Fiction   Friendship   Garrison, Jim,--1921-   History   Influence (Literary, artistic, etc.)   International relations   Johnson, Lyndon B.--1908-1973   Juvenile works   Kennedy family   Kennedy, John F.--1917-1963   Kennedy, Joseph P.--1888-1969   Kennedy, Robert F.,--1925-1968   Khrushchev, Nikita Sergeevich,--1894-1971   King, Martin Luther,--Jr.,--1929-1968   Legislators   Manners and customs   Military leadership   Nixon, Richard M.--1913-1994   Onassis, Jacqueline Kennedy,--1929-1994   Oswald, Lee Harvey   Pictorial works   Political science   Politicians   Presidents   Presidents' spouses   Presidents--Assassination   Presidents--Election   Sources   Soviet Union   United States   United States.--Central Intelligence Agency   United States.--Congress.--Senate

"LET THE WORD GO FORTH"
THE SPEECHES, STATEMENTS, AND WRITINGS OF JOHN F. KENNEDY 1947 TO 1963

November 1963
THE DEATH OF A PRESIDENT
WILLIAM MANCHESTER

PRESIDENT JOHN F KENNEDY
ASSASSINATION REPORT OF THE WARREN COMMISSION

WorldCat® Identities [Beta]

Matches for 'kennedy, david'

1. Kennedy, David published: 1854-2008

2. Kennedy, David M. published: 1939-2008

3. David M. Kennedy Center for International Studies published: 1983-2006

4. Kennedy, David W. published: 1981-2006

5. Kennedy, David (David Daniel) published: 1998-2007

6. Kennedy, David 1954-

7. Kennedy, D. L. published: 1980-2007

8. Kennedy, David M. 1958-

9. Kennedy, David 1959-

10. Kennedy, David W. 1948-

11. Kennedy, David 1943-

12. Kennedy, David 1969-

13. Kennedy, David Dr published: 1979-1994

14. Kennedy, D. B. (David Bernard) published: 1976-1996

Results Key

- A Personal Identity
- A Personal Identity from a controlled vocabulary
- A Corporate Identity
- A Corporate Identity from a controlled vocabulary

Text Size indicates relative popularity of an Identity

(Subject) The information in parentheses indicate the Subject Heading associated with that Identity

Kennedy, David 1959- [WorldCat Identities] - Windows Internet Explorer

http://www.worldcat.org/identities/lccn-nr94-35200

Google

File   Edit   View   Favorites   Tools   Help

Google

Search · · · · Bookmarks · Check · AutoFill ·

Sign In

Kennedy, David 1959- [WorldCat Identities]

Page · Tools ·

Kennedy, David 1959- [WorldCat Identities] - Windows Internet Explorer

http://www.worldcat.org/identities/lccn-nr94-35200

Google

File   Edit   View   Favorites   Tools   Help

Google

Search · · · · Bookmarks · Check · AutoFill ·

Sign In

Kennedy, David 1959- [WorldCat Identities]

Page · Tools ·

Wor

Jump To:

Kenne

Overview

W

Classificat

Publicat

1990

K

Most wid

Elegy by
2 editions pu

The new
2 editions pu

New rela

http://www.w

start

- Cornell, Joseph
- Dunn, Douglas
- Sheffield Universi
- Tuma, Keith 1957
- Pilling, Christoph
- Kennedy, Christin
- NetLibrary, Inc

Useful Links

- Library of Congress A

Associated Subj

1900 - 1999   1900

Death in literatu

ceremonies in lit

form   Little magazines

Poetry   Prose poem

© 2008 OCLC Online Computer
WorldCat Identities is covered b
OCLC 6565 Kilgour Place, Dubl

Done

start

GetRecord - Windows Internet Explorer

http://errol.oclc.org/laf/nr94-35200.html

Google

File   Edit   View   Favorites   Tools   Help

Google

Search · · · · Bookmarks · Check · AutoFill ·

GetRecord

OCLC ONLINE COMPUTER LIBRARY CENTER

A Project of OCLC Research

LAF
Linked Authority File

OCLC

Identifier:	nr94-35200
Persistent URL:	http://errol.oclc.org/laf/nr94-35200.html
XML Record:	http://errol.oclc.org/laf/nr94-35200.MarcXML

000 00037cz 2200037n 45 0
001 oca03706465
005 20070321094408.0
008 941024n| acannaab |n aaa c
010 |anr 94035200
040 |aNjP|cNjP
100 10|aKennedy, David,|d1959-
670 |aThe new poetry, c1993:|bt.p. (David Kennedy) p. 346 (b. 1959 in Leicester; senior manager in manufact. industry; poet and critic)

Log in / create account

| article | discussion | edit this page | history |

David Kennedy (disambiguation)

From Wikipedia, the free encyclopedia

David Kennedy may refer to:

- David Kennedy (1955-1984), American writer and son of politician Robert Kennedy
- David Kennedy (Australian politician) (born 1940), politician and member of the Australian House of Representatives
- David Kennedy (guitarist) (born 1976), American rock musician
- David Kennedy (jurist) (born 1954), American legal academic and scholar of international law, vice president of of international affairs at Brown University, formerly of Harvard
- David Kennedy (actor), British actor
- David Kennedy (advertising) (born 1939), American advertising executive and co-founder of Wieden+Kennedy advertising agency
- David Kennedy (film producer), producer of *Saving Milly* and the upcoming *Dark Shadows* film adaptation
- David Kennedy (New Zealand) (1864-1936), New Zealand priest, astronomer and educator.
- David M. Kennedy (1915-1996), American businessman, economist and former United States Secretary of the Treasury
- David M. Kennedy (historian), Pulitzer Prize-winning historian and professor at Stanford
- David T. Kennedy (born 1934), U.S. politician
- Dave Kennedy (born 1953), Irish racing driver

This disambiguation page lists articles associated with the same personal name. If an internal link led you here, you may wish to change the link to point directly to the intended article.

Categories: Human name disambiguation pages

WIKIPEDIA
The Free Encyclopedia

Done   Internet   100%

start   g... g... g... m... o... I... W... M... D...   11:00 AM

Research questions

- How do we establish reliable links between structured and unstructured resources?

- How do we expedite the automated processing of WorldCat Identities data?

- How do we populate new authority records?

[ORG R]) [ORG Y] Smith ([LOC D]) [ORG Y Younge] ([LOC D]) CHANGE OF SPONSORSHIP With the consent of the affected members , [ORG Representative Froehlich] was removed as principal sponsor , and [ORG Representative Feigenholtz] became the new principal sponsor of HOUSE BILL 4995 . With the consent of the affected members , [ORG Representative Soto] was removed as principal sponsor , and [ORG Representative Ryg] became the new principal sponsor of SENATE BILL 222 . HOUSE RESOLUTION The following resolution was offered and placed in the [ORG Committee] on Rules . HOUSE RESOLUTION 1009 Offered by Representative [PER Madigan:] [LOC WHEREAS] , [PER FutureGen

---

article   discussion   edit this page   history

FutureGen

From Wikipedia, the free encyclopedia
(Redirected from Futuregen)

This article or section contains information about a proposed, planned construction power station.

It may contain information of a speculative nature and the content may change as the construction completion of the power station approaches, and more information becomes available.

FutureGen is a US government project announced by President George W. Bush in 2003[1]; its initial plan involved the construction of a near zero-emissions coal-fueled power plant to produce hydrogen and electricity while using carbon capture and storage.[2]

In December 2007, Mattoon Township, Coles County, Illinois northwest of Mattoon, Illinois was chosen as the site for the plant from among four finalists in Illinois and Texas. On

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

search

[Go]   [Search]

..and an opportunity

- Use WorldCat Identities as a resource for establishing the referent of a name

- Build on research at UIUC that uses Wikipedia for identity resolution

- Utilize domain specific proper name resources (e.g. Union List of Artist Names ULAN from Getty)

Matching names to identities

Extract Names from Text + **Identifying information**

David Kennedy

Kennedy

born...1959—Leicester, England

The New Poetry

Search WorldCat Identities

Kennedy, David 1954-

Kennedy, D. L. published: 1980-2007

Kennedy, David M. 1958-

Kennedy, David 1959-

Kennedy, David W. 1948-

Kennedy, David 1943-

Match identity info from text

And Identities pages

A unique identifier

John F. Kennedy

Extract Names from Text + **Identifying information**

John F. Kennedy

Kennedy, JFK

Married to Jacqueline Lee Beauvier

Search WorldCat Identities

Kennedy, John F. published: 1940-2008 (Children of presidents--United States) Kennedy, John F. (John Fitzgerald) 1917-1963 (Presidents--United States) Kennedy, John F. (John Fitzgerald) (Children of presidents--United States) John F. Kennedy Center for the Performing Arts (U.S.) published: 1959-2008

Match identity info from text

And Identities pages

A unique identifier

Project status

- Evaluating annotated data
- Developing workflows
- Demo building
- Integrating NER tool with CLiMB
- Proof-of-concept results of identity resolution algorithms
- Building partners

References

- NDIIPP site: http://ndiipp.uiuc.edu/
- CLiMB (Computational Linguistics for Metadata Building) site: http://www.umiacs.umd.edu/~climb/
- Named Entity Recognizer (NER) UIUC homepage: http://l2r.cs.uiuc.edu/
- OCLC's WorldCat Identities: http://www.worldcat.org/identities/

EMP Project Team - 1

- Jean Godby, Project Lead, OCLC
  - godby@oclc.org
- Judith Klavans, Project Lead, UMD
  - jklavans@umd.edu
- Dan Roth, Project Lead, UIUC
  - danr@illinois.edu
- Patricia Hswe, NDIIPP Project Manager, UIUC – phswe@illinois.edu

EMP Project Team - 2

- Sarah Dotson, User Studies, UMD
  - sjdotson@gmail.com
- Larry Jackson, Collection Consultant, UIUC – lsjackso@illinois.edu
- Rebecca LaPlante, Project Coordinator, UMD – laplante@umd.edu
- Lev-Arie Ratinov, Graduate Research Assistant, UIUC - ratinov2@uiuc.edu
- Devon Smith, Software Engineer, OCLC
  - smithde@oclc.org

Summary

Thank you.

Questions?

Notation and Task Definition

- 'Mention' (def) a string in text that refers to a real-world entity; 'Entity' (def) is the entity being referred to.

- The goal: given unlabeled text, link mentions to entities. ...

Document 1: *The Justice Department has officially ended its inquiry into the assassinations of* **John F. Kennedy** *and Martin Luther King Jr., finding ``no persuasive evidence'' to support conspiracy theories, according to department documents. The House Assassinations Committee concluded in 1978 that* **Kennedy** *was ``probably'' assassinated as the result of a conspiracy involving a second gunman, a finding that broke from the* **Warren Commission's** *belief that Lee Harvey Oswald acted alone in* **Dallas** *on Nov. 22, 1963.*

Document 2: **David Kennedy** *was born in Leicester, England in 1959.* *...***Kennedy** *co-edited The New Poetry (Bloodaxe Books 1993), and is the author of New Relations: The Refashioning Of British Poetry 1980-1994 (Seren 1996).*