S3K
Seeking Statement-Supporting top-K Witnesses

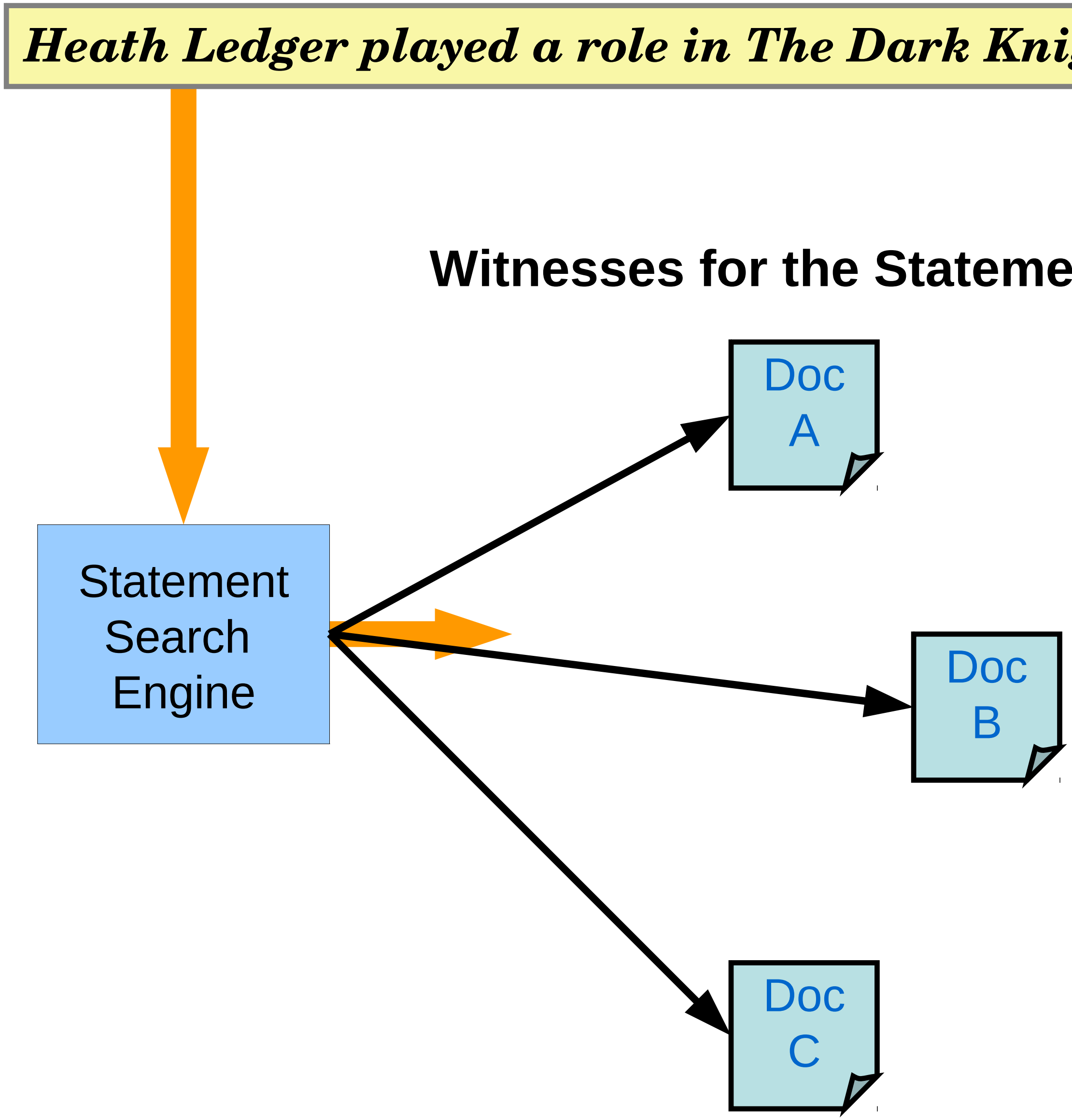
Steffen Metzger*, Shady Elbassuoni*,
Katja Hose*, Ralf Schenkel*+
*Max-Planck-Institute for Informatics
+Saarland University

CIKM 2011
25th October 2011

Statements

- Statements = „factual statements“
  - Obama was born in Hawaii
  - Obama was born in Kenya
  - Heath Ledger played a role in The Dark Knight
  - Maggie Gyllenhaal appeared in The Dark Knight
  - …
- „statement“ used interchangeably with „fact“, but no truthness assumed in general

Document Search by Statements

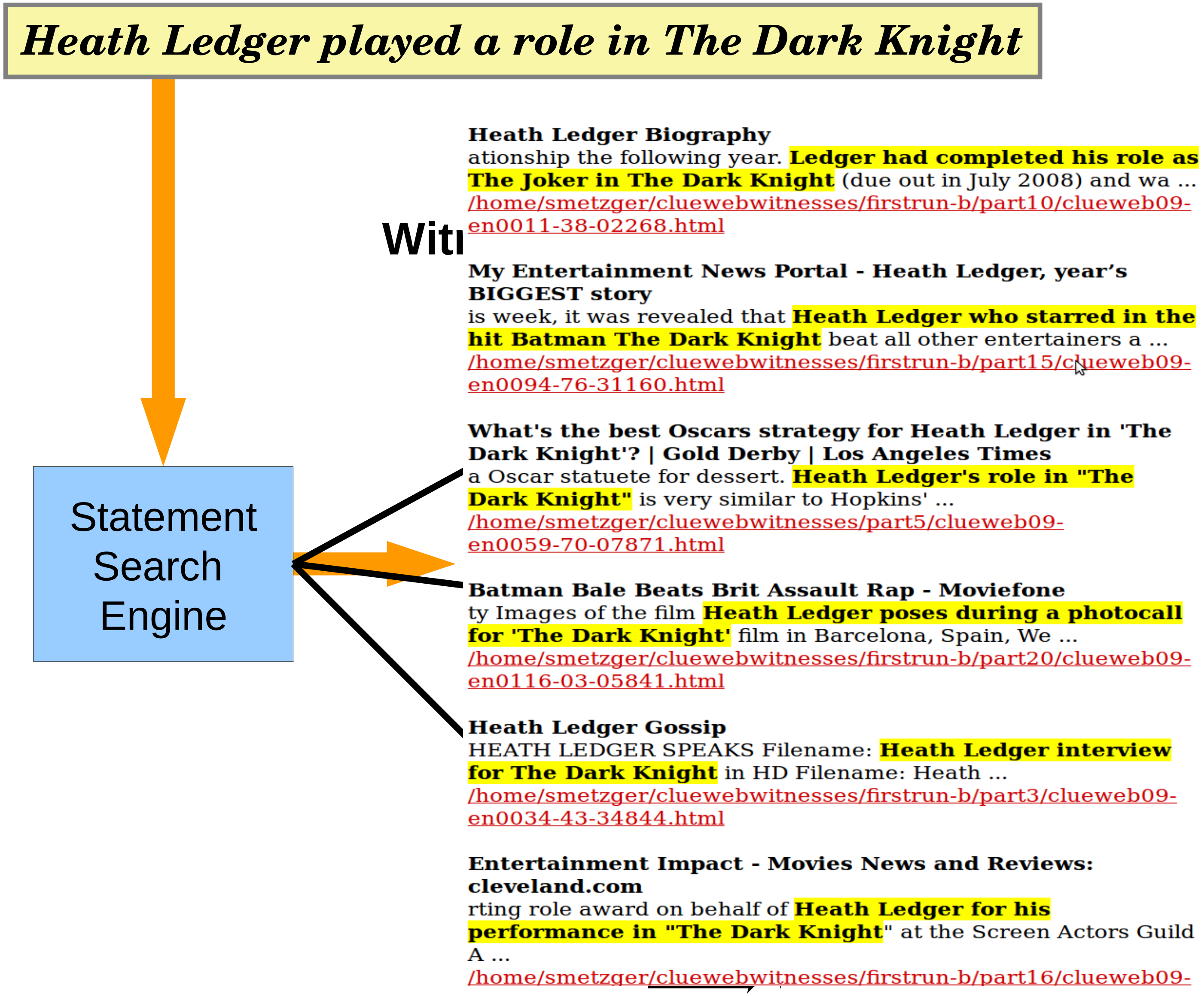


Heath Ledger played a role in The Dark Knight

Witnesses for the Statement(s)

Statement Search Engine

Document Search by Statements



Heath Ledger played a role in The Dark Knight

Statement Search Engine

Wit

Heath Ledger Biography
atîonship the following year. **Ledger had completed his role as The Joker in The Dark Knight** (due out in July 2008) and wa ...
/home/smetzger/cluewebwitnesses/firstrun-b/part10/clueweb09-en0011-38-02268.html

My Entertainment News Portal - Heath Ledger, year’s BIGGEST story
is week, it was revealed that **Heath Ledger who starred in the hit Batman The Dark Knight** beat all other entertainers a ...
/home/smetzger/cluewebwitnesses/firstrun-b/part15/clueweb09-en0094-76-31160.html

What's the best Oscars strategy for Heath Ledger in 'The Dark Knight'? | Gold Derby | Los Angeles Times
a Oscar statuete for dessert. **Heath Ledger's role in "The Dark Knight"** is very similar to Hopkins' ...
/home/smetzger/cluewebwitnesses/part5/clueweb09-en0059-70-07871.html

Batman Bale Beats Brit Assault Rap - Moviefone
ty Images of the film **Heath Ledger poses during a photocall for 'The Dark Knight'** film in Barcelona, Spain, We ...
/home/smetzger/cluewebwitnesses/firstrun-b/part20/clueweb09-en0116-03-05841.html

Heath Ledger Gossip
HEATH LEDGER SPEAKS Filename: **Heath Ledger interview for The Dark Knight** in HD Filename: Heath ...
/home/smetzger/cluewebwitnesses/firstrun-b/part3/clueweb09-en0034-43-34844.html

Entertainment Impact - Movies News and Reviews: cleveland.com
rting role award on behalf of **Heath Ledger for his performance in "The Dark Knight"** at the Screen Actors Guild A ...
/home/smetzger/cluewebwitnesses/firstrun-b/part16/clueweb09-

Statement Expressions

- One statement, many forms of verbal expression

Ledger's performance in The Dark Knight...

...Ledger's latest movie The Dark Knight...

Heath was great in The Dark Knight...

...Ledger playing the Joker in The Dark Knight...

Statement Expressions

- One statement, many forms of verbal expression

Ledger's performanc .

Did Heath Ledger play a role in The Dark Knight?

...Ledger's latest movie The Dark Knight...

Ledger acted in the new Batman movie?

Heath was great in The Dark Knight...

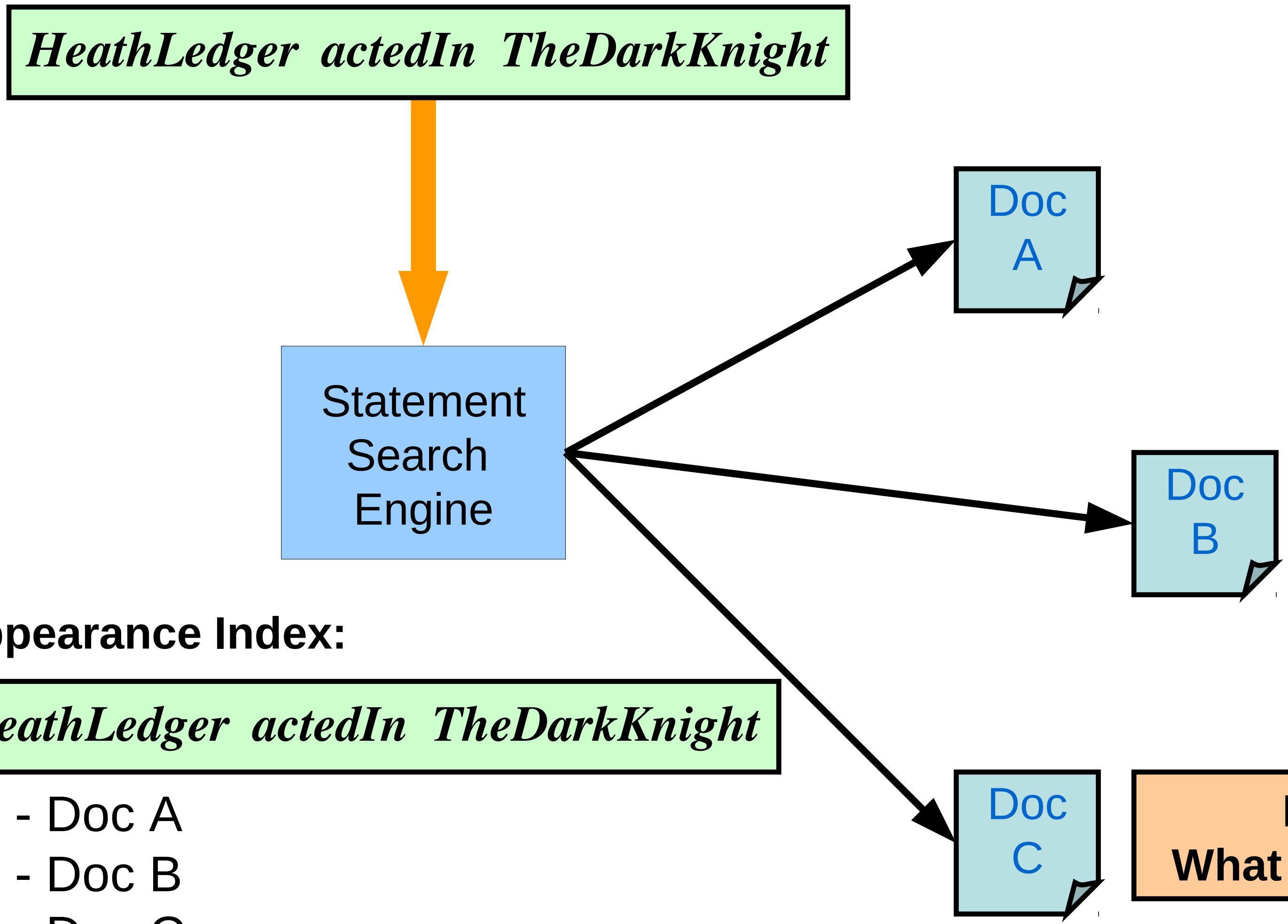
playing the Joker in The Dark Knight...

- **Unify into** single representation: **RDF triples**

HeathLedger actedIn TheDarkKnight

Subject Relation Object

Document Search by Statements



HeathLedger actedIn TheDarkKnight

Statement Search Engine

Doc A

Doc B

Doc C

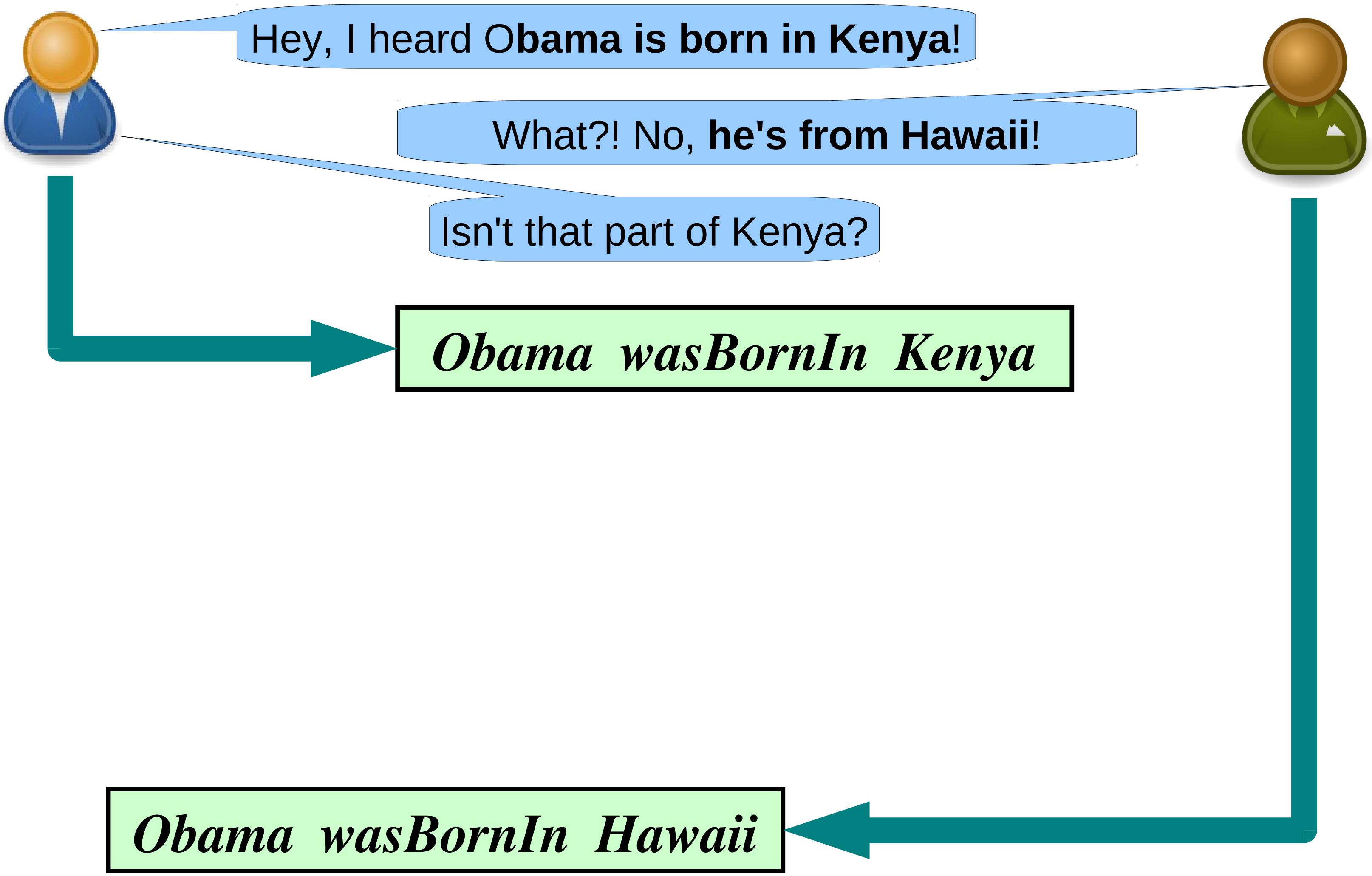
Appearance Index:

HeathLedger actedIn TheDarkKnight

- Doc A
- Doc B
- Doc C

But wait!
What is it good for?

Information Need: Verification



Hey, I heard **Obama is born in Kenya**!

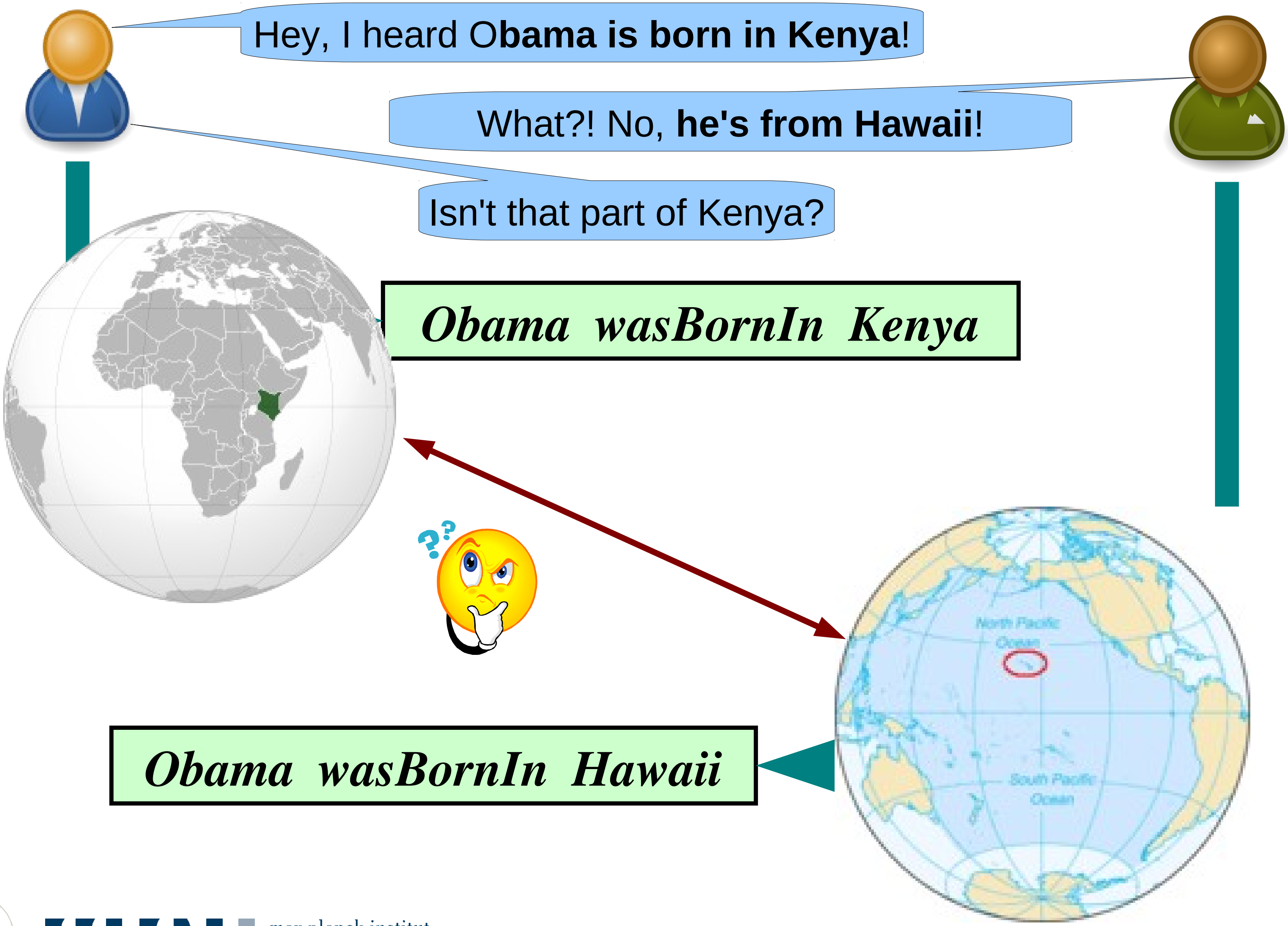
What?! No, **he's from Hawaii**!

Isn't that part of Kenya?

Obama wasBornIn Kenya

Obama wasBornIn Hawaii

Information Need: Verification



Hey, I heard **Obama is born in Kenya**!

What?! No, **he's from Hawaii**!

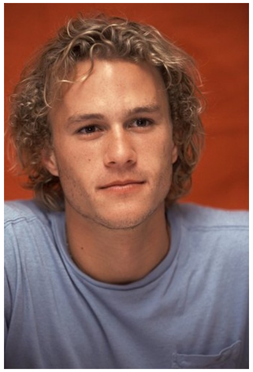
Isn't that part of Kenya?

Obama wasBornIn Kenya

Obama wasBornIn Hawaii

Information Need: Context

HeathLedger actedIn TheDarkKnight



MaggieGyllenhaal actedIn TheDarkKnight

Did they have an affair?

What role did he play?

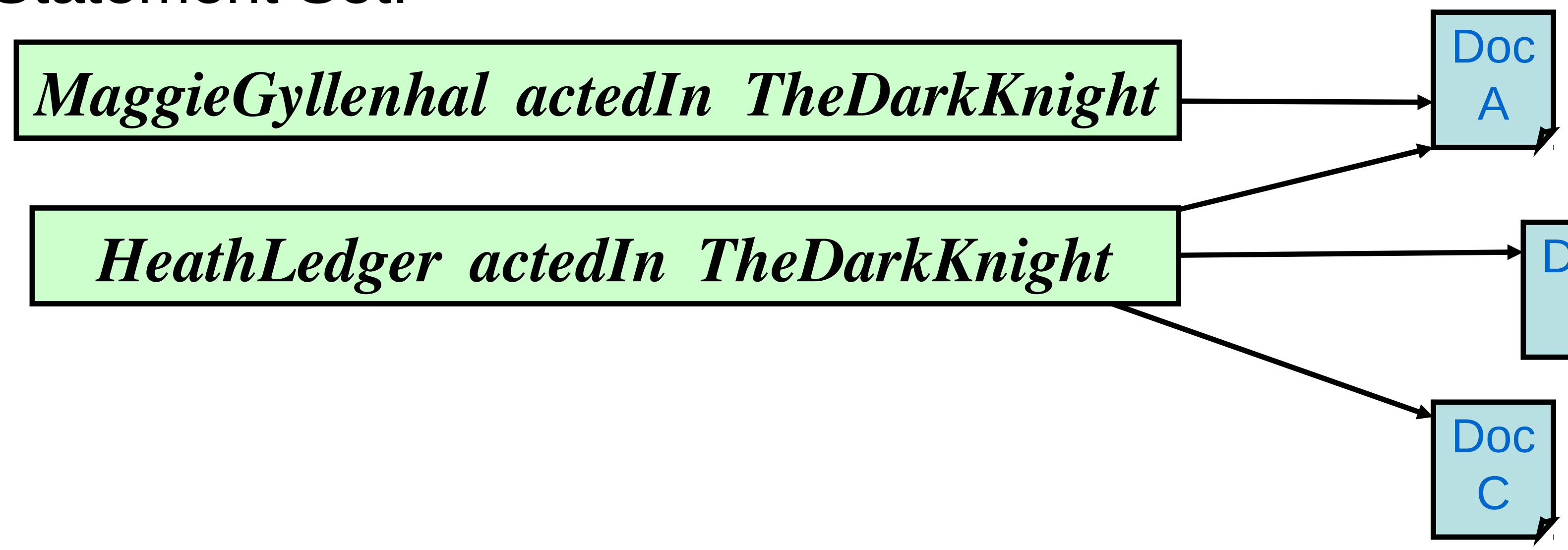
What's the relation of their characters?

What's his interpretation of that role?

Ranked Witness List

Ranked list of source documents

Statement Set:


MaggieGyllenhal actedIn TheDarkKnight
HeathLedger actedIn TheDarkKnight
Doc A
Doc B
Doc C

How should we rank?

Not every witness is equally appropriate

Goooooaaaal!

Not every witness is equally appropriate

Some witnesses have more Authority than others (witness trust)

Testimonies are not equally convincing

Obama visited his home country Kenya.

Barack Obama was born in Honolulu, Hawaii in 1961

Ledger's latest movie The Dark Knight topped the box office ...

Ledger playing the Joker in The Dark Knight...

Testimonies are not equally convincing

Obama visited his home country Kenya.

Barack Obama was born in Honolulu, Hawaii in 1961

Ledger's latest movie The Dark Knight topped the box office ...

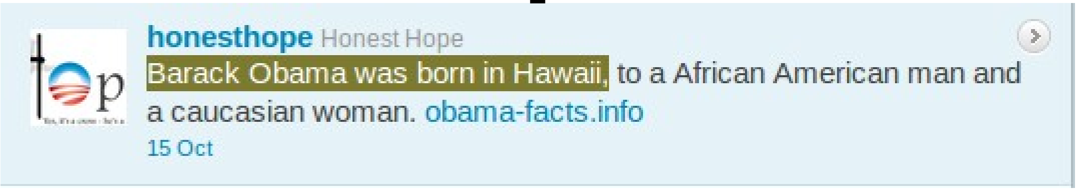
I

Persuasiveness of testimonies varies (strength of statement support)

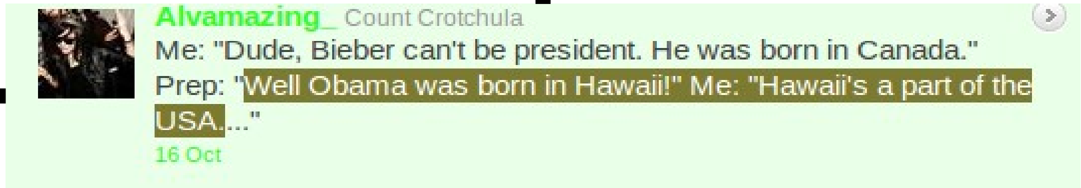
The more the better

Obama wasBornIn Hawaii

Hawaii isPartOf USA



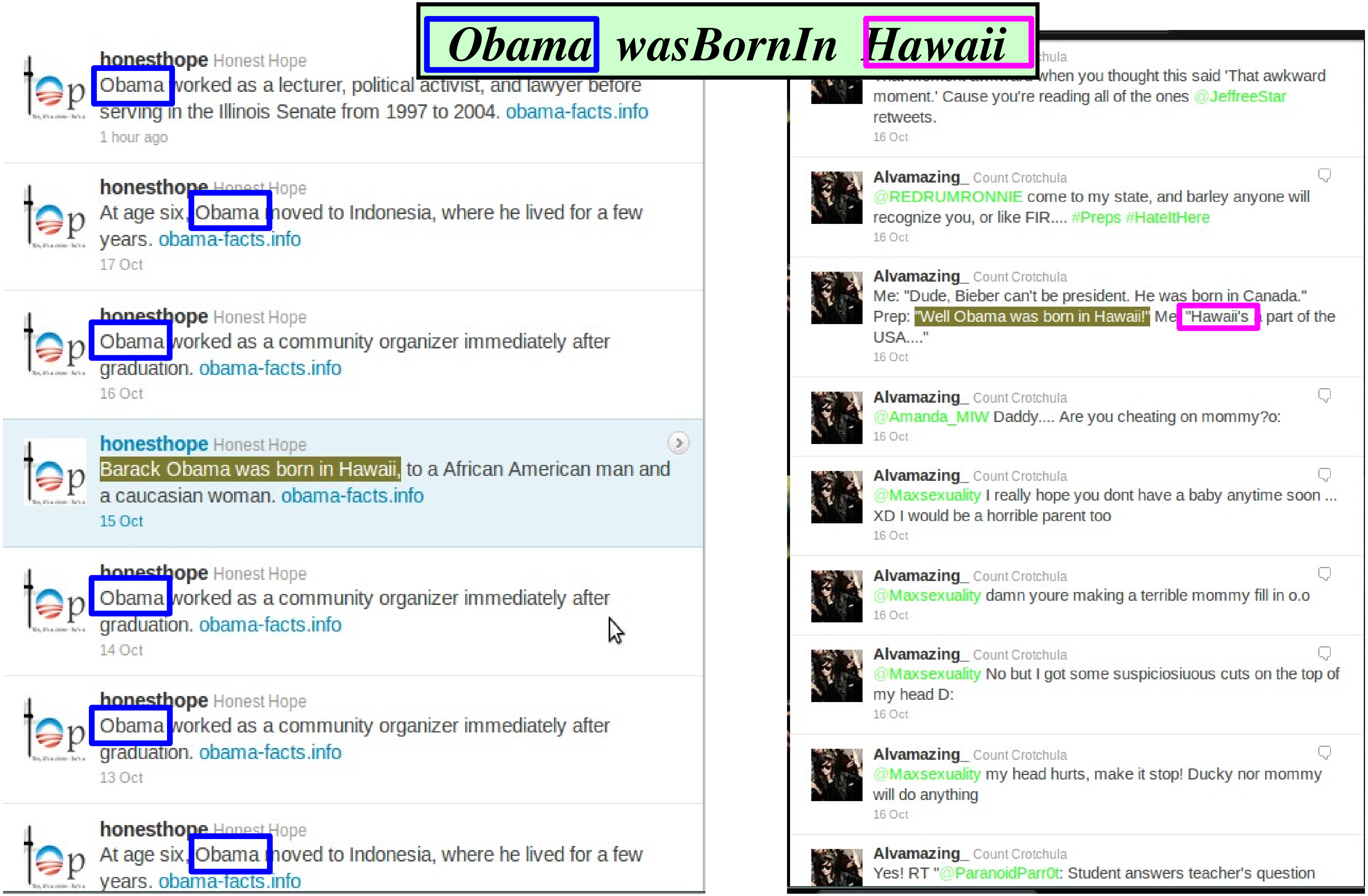
honesthope Honest Hope
Barack Obama was born in Hawaii, to a African American man and a caucasian woman. obama-facts.info
15 Oct



Alvamazing_ Count Crotchula
Me: "Dude, Bieber can't be president. He was born in Canada." Prep: "Well Obama was born in Hawaii!" Me: "Hawaii's a part of the USA...."
16 Oct

The more statements covered the better (coverage)

The more the better



Obama wasBornIn Hawaii

honesthope Honest Hope
Obama worked as a lecturer, political activist, and lawyer before serving in the Illinois Senate from 1997 to 2004. obama-facts.info
1 hour ago

honesthope Honest Hope
At age six, Obama moved to Indonesia, where he lived for a few years. obama-facts.info
17 Oct

honesthope Honest Hope
Obama worked as a community organizer immediately after graduation. obama-facts.info
16 Oct

honesthope Honest Hope
Barack Obama was born in Hawaii, to a African American man and a caucasian woman. obama-facts.info
15 Oct

honesthope Honest Hope
Obama worked as a community organizer immediately after graduation. obama-facts.info
14 Oct

honesthope Honest Hope
Obama worked as a community organizer immediately after graduation. obama-facts.info
13 Oct

honesthope Honest Hope
At age six, Obama moved to Indonesia, where he lived for a few years. obama-facts.info

when you thought this said 'That awkward moment.' Cause you're reading all of the ones @JeffreeStar retweets.
16 Oct

Alvamazing_ Count Crotchula
@REDRUMRONNIE come to my state, and barley anyone will recognize you, or like FIR.... #Preps #HateItHere
16 Oct

Alvamazing_ Count Crotchula
Me: "Dude, Bieber can't be president. He was born in Canada." Prep: "Well Obama was born in Hawaii!" Me: "Hawaii's a part of the USA...."
16 Oct

Alvamazing_ Count Crotchula
@Amanda_MIW Daddy.... Are you cheating on mommy?o:
16 Oct

Alvamazing_ Count Crotchula
@Maxsexuality I really hope you dont have a baby anytime soon ... XD I would be a horrible parent too
16 Oct

Alvamazing_ Count Crotchula
@Maxsexuality damn youre making a terrible mommy fill in o.o
16 Oct

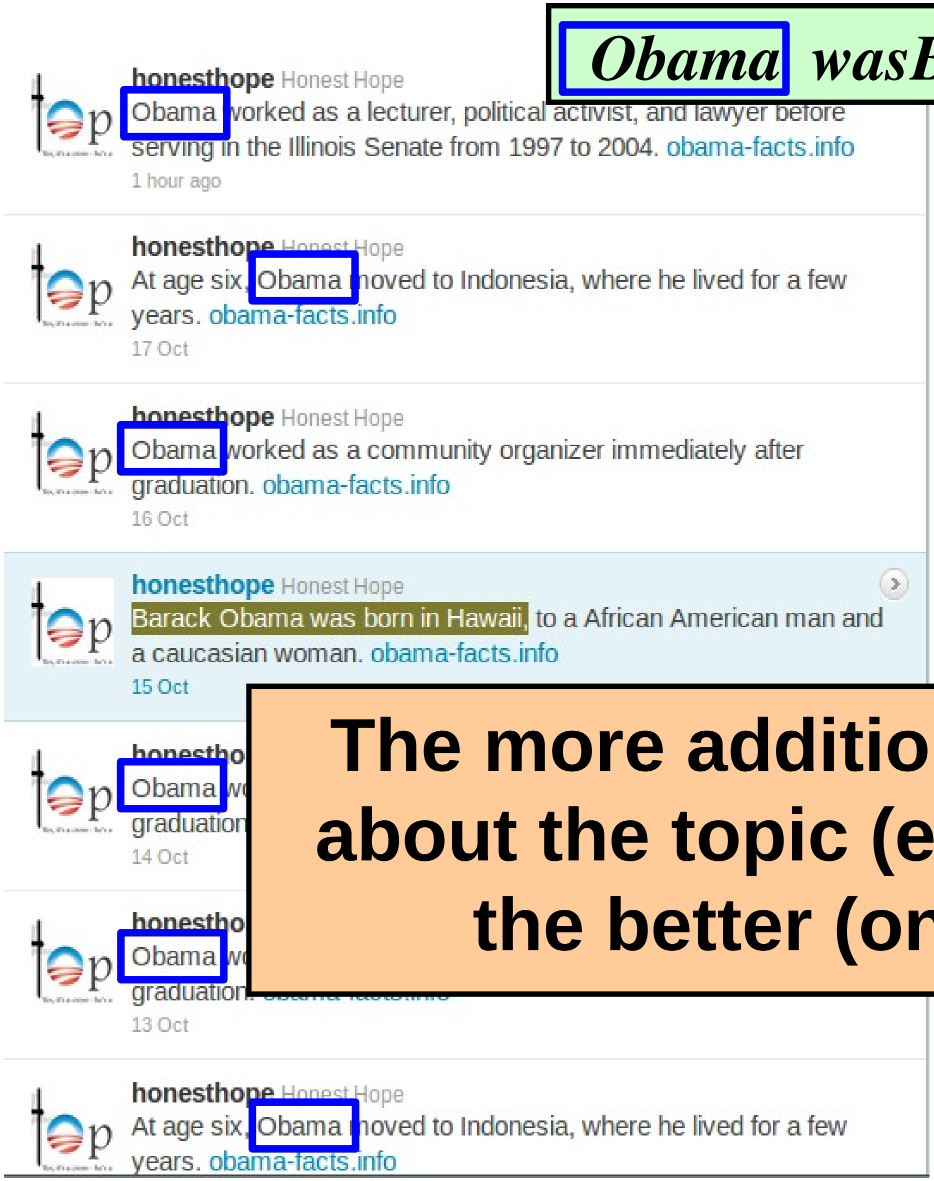
Alvamazing_ Count Crotchula
@Maxsexuality No but I got some suspiciosiuous cuts on the top of my head D:
16 Oct

Alvamazing_ Count Crotchula
@Maxsexuality my head hurts, make it stop! Ducky nor mommy will do anything
16 Oct

Alvamazing_ Count Crotchula
Yes! RT "@ParanoidParr0t: Student answers teacher's question

The more the better

Obama *wasBornIn* *Hawaii*



honesthope Honest Hope
Obama worked as a lecturer, political activist, and lawyer before serving in the Illinois Senate from 1997 to 2004. obama-facts.info
1 hour ago

honesthope Honest Hope
At age six, Obama moved to Indonesia, where he lived for a few years. obama-facts.info
17 Oct

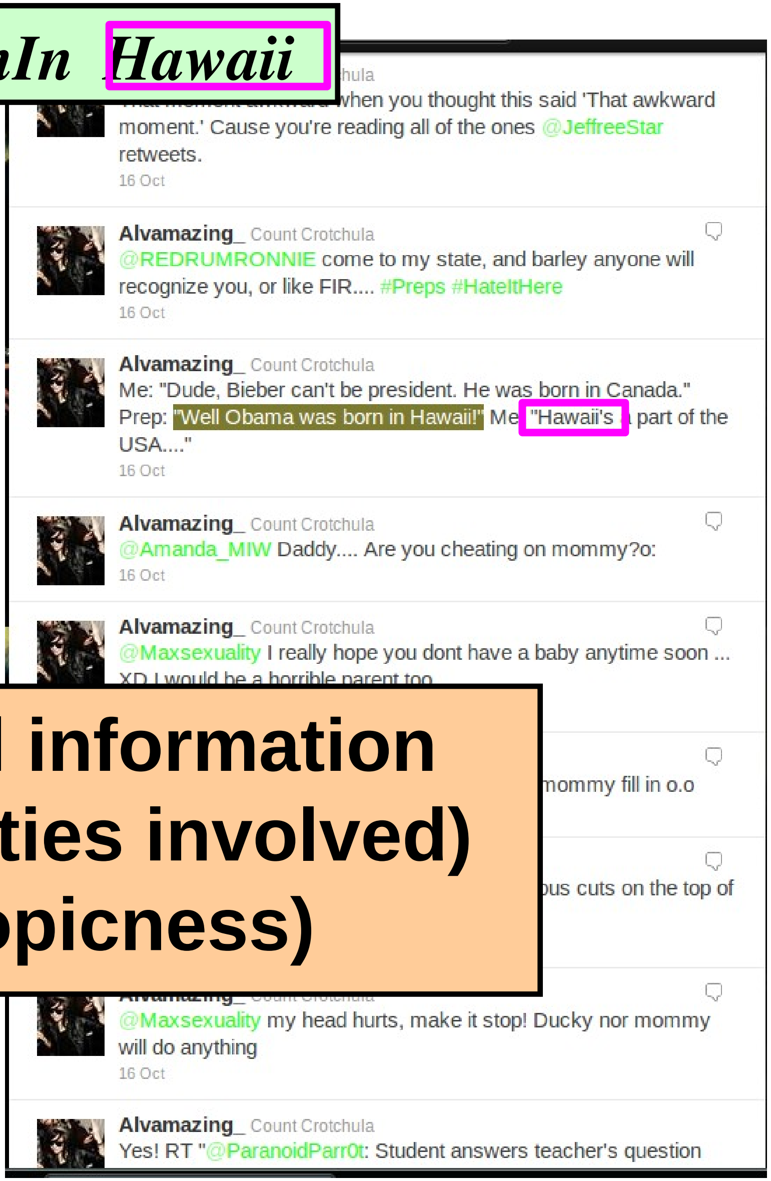
honesthope Honest Hope
Obama worked as a community organizer immediately after graduation. obama-facts.info
16 Oct

honesthope Honest Hope
Barack Obama was born in Hawaii, to a African American man and a caucasian woman. obama-facts.info
15 Oct

honestho
Obama w
graduation
14 Oct

honestho
Obama w
graduation.
13 Oct

honesthope Honest Hope
At age six, Obama moved to Indonesia, where he lived for a few years. obama-facts.info



hula
when you thought this said 'That awkward moment.' Cause you're reading all of the ones @JeffreeStar retweets.
16 Oct

Alvamazing_ Count Crotchula
@REDRUMRONNIE come to my state, and barley anyone will recognize you, or like FIR.... #Preps #HateItHere
16 Oct

Alvamazing_ Count Crotchula
Me: "Dude, Bieber can't be president. He was born in Canada." Prep: "Well Obama was born in Hawaii!" Me: "Hawaii's a part of the USA...."
16 Oct

Alvamazing_ Count Crotchula
@Amanda_MIW Daddy.... Are you cheating on mommy?o:
16 Oct

Alvamazing_ Count Crotchula
@Maxsexuality I really hope you dont have a baby anytime soon ... XD I would be a horrible parent too

mommy fill in o.o

us cuts on the top of

@Maxsexuality my head hurts, make it stop! Ducky nor mommy will do anything
16 Oct

Alvamazing_ Count Crotchula
Yes! RT "@ParanoidParr0t: Student answers teacher's question

The more additional information about the topic (entities involved) the better (on-topicness)

Criteria for ranking

- **Coverage** – how many of the statements are covered in the document?
- **Persuasiveness** – how convincing are the sources
  - **Authority** of the document
  - **Strength** of statement support
- **On-Topicness** – is the focus of the document on the topic expressed by the statements?

Criteria for ranking

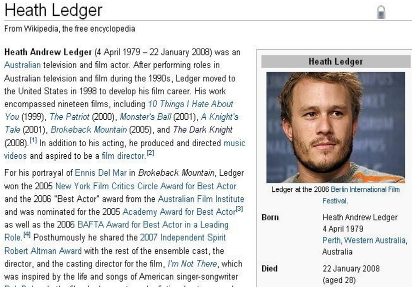
- **Coverage** – how many of the statements are covered in the docu
- **Persuasiveness** – h sources
  - – **Authority** of the document
  - – **Strength** of statement support
- **On-Topicness** – is the focus of the document on the topic (ements?

1) How to connect documents with statements?

2) And how can we measure a document's persuasiveness?

„Easy“: Appearances of the entities occurring in the statement(s)

Leveraging IE Methods



Heath Ledger

From Wikipedia, the free encyclopedia

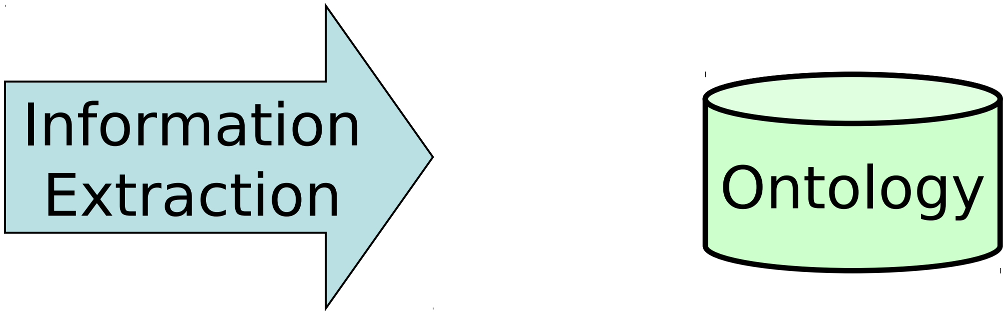
Heath Andrew Ledger (4 April 1979 – 22 January 2008) was an Australian television and film actor. After performing roles in Australian television and film during the 1990s, Ledger moved to the United States in 1998 to develop his film career. His work encompassed nineteen films, including *10 Things I Hate About You* (1999), *The Patriot* (2000), *Monster's Ball* (2001), *A Knight's Tale* (2001), *Brokeback Mountain* (2005), and *The Dark Knight* (2008).[1] In addition to his acting, he produced and directed music videos and aspired to be a film director.[2]

For his portrayal of Ennis Del Mar in *Brokeback Mountain*, Ledger won the 2005 New York Film Critics Circle Award for Best Actor and the 2006 "Best Actor" award from the Australian Film Institute and was nominated for the 2005 Academy Award for Best Actor[3] as well as the 2006 BAFTA Award for Best Actor in a Leading Role.[4] Posthumously he shared the 2007 Independent Spirit Robert Altman Award with the rest of the ensemble cast, the director, and the casting director for the film, *I'm Not There*, which was inspired by the life and songs of American singer-songwriter

Heath Ledger

Ledger at the 2006 Berlin International Film Festival.

Born	Heath Andrew Ledger 4 April 1979 Perth, Western Australia, Australia
Died	22 January 2008 (aged 28)



Information Extraction

Ontology

Heath Ledger was born in Perth, Australia...

Pattern *p1*:
X was born in Y

HeathLedger wasBornIn Perth(Australia)

...Ledger's latest movie The Dark Knight...

Pattern *p2*:
X 's latest movie Y

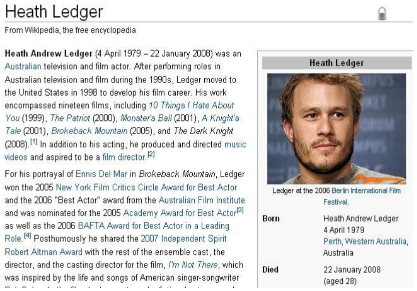
HeathLedger actedIn TheDarkKnight

Information

Extraction Rules,
Patterns

Knowledge

Leveraging IE Methods



Heath Ledger

From Wikipedia, the free encyclopedia

Heath Andrew Ledger (4 April 1979 – 22 January 2008) was an Australian television and film actor. After performing roles in Australian television and film during the 1990s, Ledger moved to the United States in 1998 to develop his film career. His work encompassed nineteen films, including *10 Things I Hate About You* (1999), *The Patriot* (2000), *Monster's Ball* (2001), *A Knight's Tale* (2001), *Brokeback Mountain* (2005), and *The Dark Knight* (2008).[1] In addition to his acting, he produced and directed music videos and aspired to be a film director.[2]

For his portrayal of Ennis Del Mar in *Brokeback Mountain*, Ledger won the 2005 New York Film Critics Circle Award for Best Actor and the 2006 "Best Actor" award from the Australian Film Institute and was nominated for the 2005 Academy Award for Best Actor[3] as well as the 2006 BAFTA Award for Best Actor in a Leading Role.[4] Posthumously he shared the 2007 Independent Spirit Robert Altman Award with the rest of the ensemble cast, the director, and the casting director for the film, *I'm Not There*, which was inspired by the life and songs of American singer-songwriter

Heath Ledger

Ledger at the 2006 Berlin International Film Festival.

Born	Heath Andrew Ledger 4 April 1979 Perth, Western Australia, Australia
Died	22 January 2008 (aged 28)

Information Extraction

Ontology

Heath Ledger was born in Perth, Australia...

Pattern *p1*:
X was born in Y

HeathLedger wasBornIn Perth(Australia)

...Ledger's latest movie The Dark Knight...

Pattern *p2*:
X 's latest movie Y

HeathLedger actedIn TheDarkKnight

Pattern reliability differs
→ Confidence values

p1 expresses wasBornIn : 0.9

p2 expresses actedIn : 0.4

p2 expresses directed : 0.5

Witness Ranking: Trust

- Rank of witness *d* based on $P(isPerfect(d,g))$
  - Probability to be a perfect witness of a given set of statements $g=\{f_1,\ldots,f_n\}$ with $f_i=(s_i,r_i,o_i)$

Witness Ranking: Trust

- Rank of witness *d* based on $P(isPerfect(d,g))$
  - Probability to be a perfect witness of a given set of statements $g=\{f_1,\ldots,f_n\}$ with $f_i=(s_i,r_i,o_i)$

$$P(isPerfect(d,g)) \propto P(trust(d))P(g|d)$$

trust estimated by pagerank *pr(d)*

$$P(isPerfect(d,g)) \propto pr(d)P(g|d)$$

Independence Collection (Col) based Smoothing

$$P(g|d)=\prod_{i=1}^{n} \alpha P(f_i|d)+(1-\alpha)P(f_i|Col)$$

Support and On-Topicness

$f_i = (s_i, r_i, o_i)$

$$P(g|d) = \prod_{i=1}^{n} \alpha P(f_i|d) + (1-\alpha) P(f_i|Col)$$

$X \in \{d, Col\}$

$$P(f_i|X) = \beta_s P_e(s_i|X) + \beta_o P_e(o_i|X) + (1-\beta_s-\beta_o) P_f(f_i|X)$$

Pattern instance with s_i, o_i

$$P_e(s|d) = \frac{count(s,d)}{\sum_{e \in d} count(e,d)}$$

$$P_f(f_i|X) = \sum_{j=0}^{m} (P(p_j^{s_i,o_i}|X) P(r_i|p_j))$$

Pattern confidence
$conf(p_j, r_i)$

$$P_f(f_i|X) = \sum_{j=0}^{m} (P(p_j^{s_i,o_i}|X) conf(p_j, r_i))$$

Statement Presence Estimation

$f_i = (s_i, r_i, o_i)$

$$P(f_i|X) = \beta_s P_e(s_i|X) + \beta_o P_e(o_i|X) + (1 - \beta_s - \beta_o) P_f(f_i|X)$$

$X \in \{d, Col\}$

$$P_f(f_i|X) = \sum_{j=0}^{m} P(p_j^{s_i, o_i}|X) conf(p_j, r_i)$$

Pattern Importance Estimation

$$P(p_j^{s_i, o_i}|X) = \frac{count(p_j^{s_i, o_i}, X)}{\sum_p count(p, X)}$$

Evaluation

- Documents: ~180.000 from *ClueWeb09*+
- IE Parser: Modified SOFIE/PROSPERA system
- Evaluation 1:
  - Comparing different ranking approaches on prefiltered witness set
    - **naive**: only filtering
    - **topic**: mainly enity occurrences
    - **persuade**: aiming for persuasiveness
    - **mix**: balancing both aims
    - **lucene**: keyword search by entities on filtered subset

+http://lemurproject.org/clueweb09.php/

Evaluation

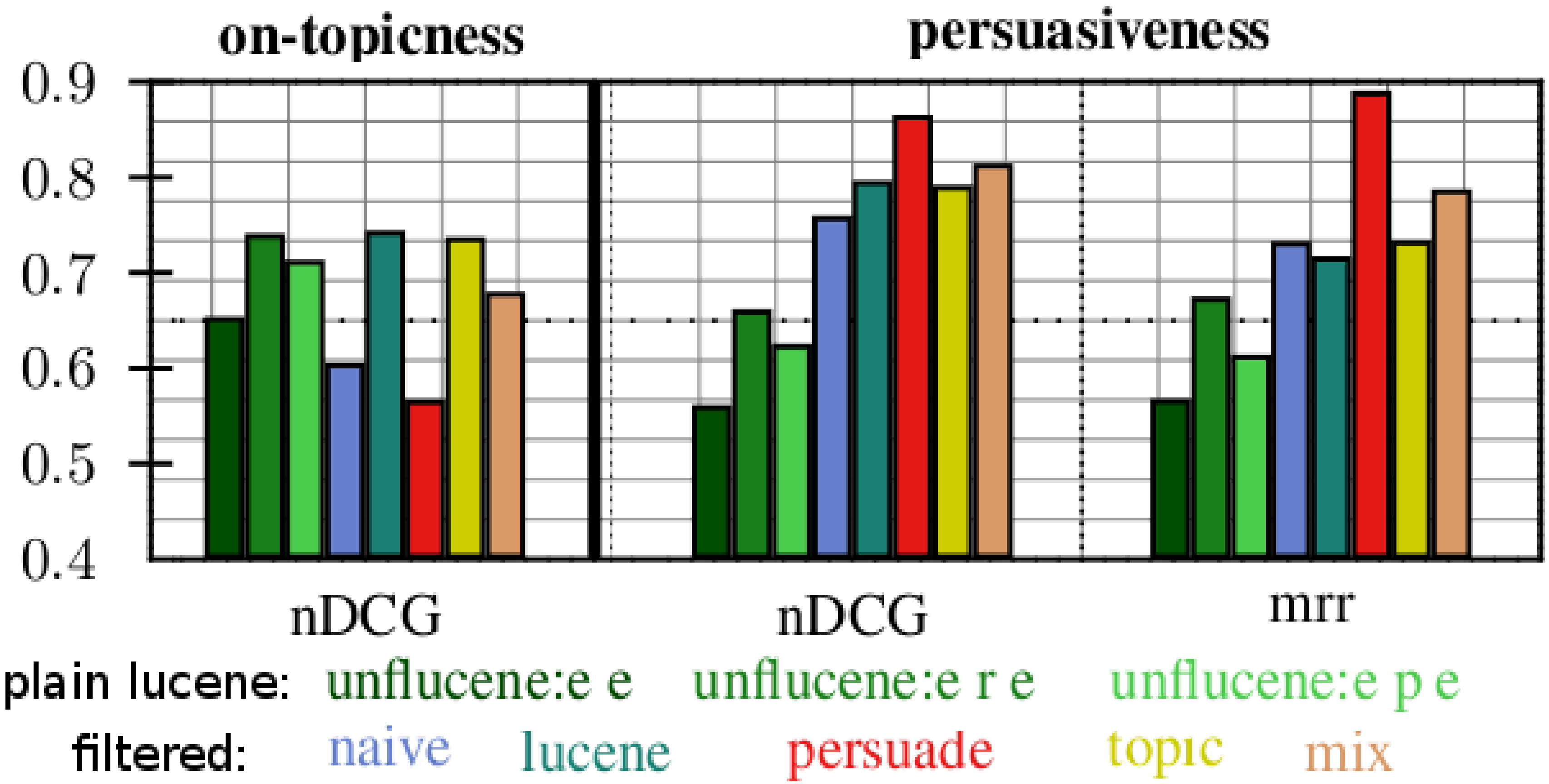
- Evaluation 2:
  - Comparing extraction based methods against plain keyword search on full document corpus
    - **unfluc: e e**
      - based on entities
    - **unfluc: e r e**
      - based on entities & manual relation formulation
    - **unfluc: e p e**
      - based on entities & best pattern for relation (highest confidence)

Evaluation

- 56 queries
  - witness pooling from various rankings (top-50, top-20)
- Graded Human Assessments
  - *On-topicness:* focus on entities of statements
  - *Persuasiveness:* how well does the document convince me of the statement(s)
  - 3 grades on both scales

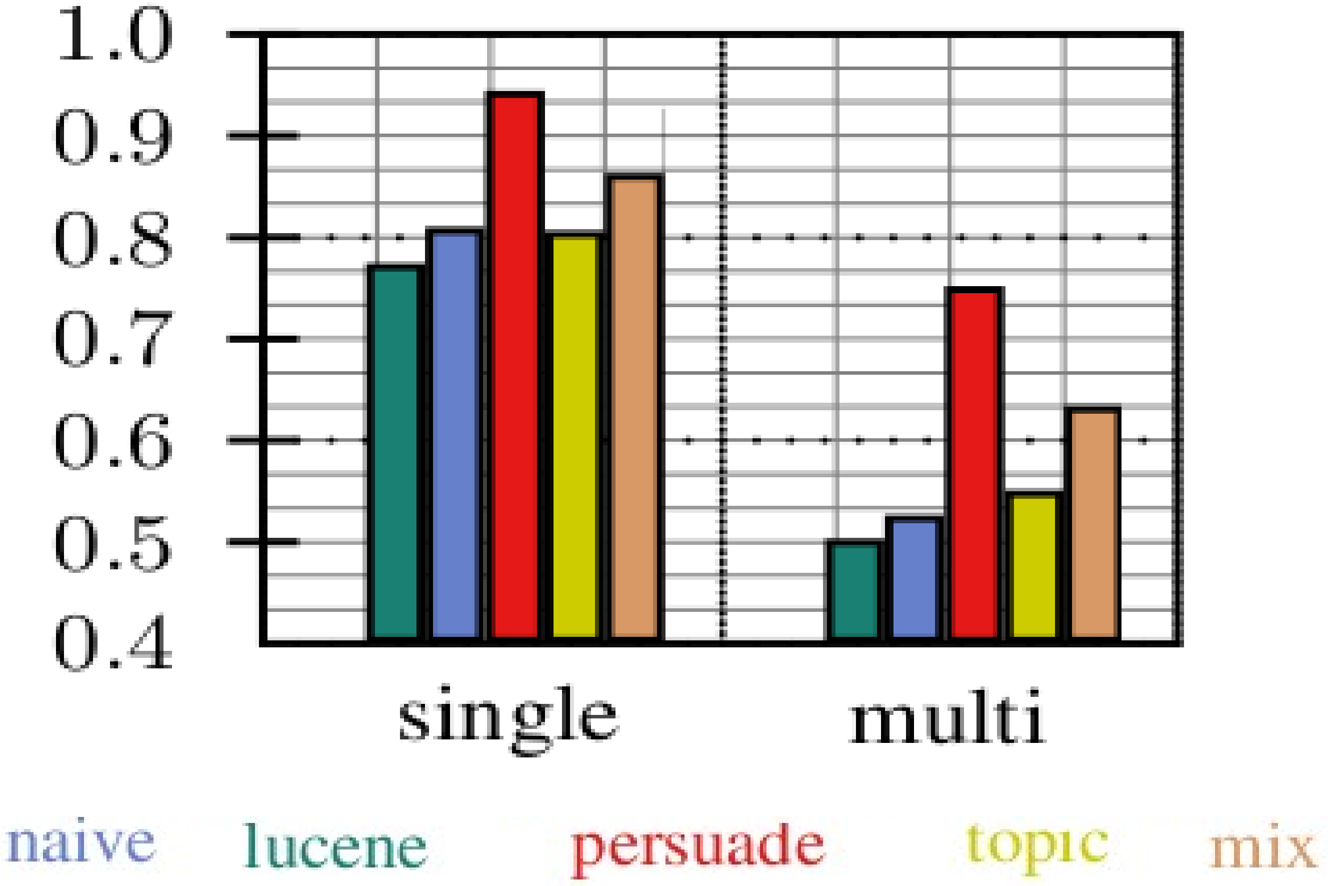
$^{+}$http://lemurproject.org/clueweb09.php/

Evaluation



Overall comparison over all queries

Evaluation – Single vs. Multi-Statement



Comparing **single-statement vs. multi-statement Persuasiveness** performance by **mrr**

Thanks for your attention

- Any questions?

Image sources used

- Mostly public domain, some CC licenses, some from free collections/fun pages assuming public domain or a fair use based usage
- Wikimedia Commons
- Cliparts (opencliparts.org, MS ClipArt Collection)
- Heath Ledger image from „Doctor Hyde“ @ Flickr, CC attribution: http://www.flickr.com/photos/indieflickr/2214583962/
- Charts generated with latex and pstricks
- Football pictures from random fun pages, no license information found. If you hold the rights on any of them and you feel infringed on your rights, please contact me.