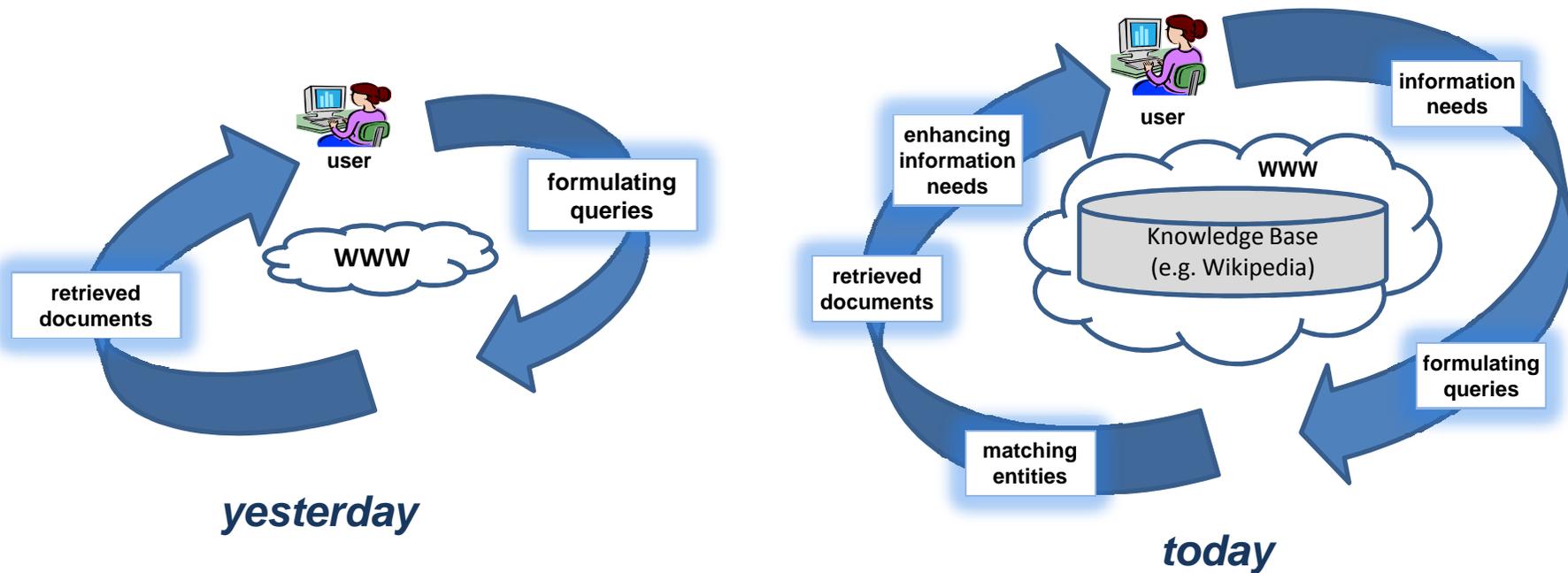


Connecting Knowledge for A New Kind of Search

Andias Wira Alam
Email: andias.wira-alam@gesis.org

Motivation and Background

Web Search



Example on Query Answering

Who is the 44th President of the USA?



Retrieved Documents:



[President Barack Obama | The White House](#)

www.whitehouse.gov > [The Administration](#) ▾ [Diese Seite übersetzen](#)

Barack H. Obama is the **44th President of the United States**. His story is the American story — values from the heartland, a middle-class upbringing in a strong ...

- The answer lies in the snippet (from one document)
- What if the answer lies in more documents?

Query Results (in general)

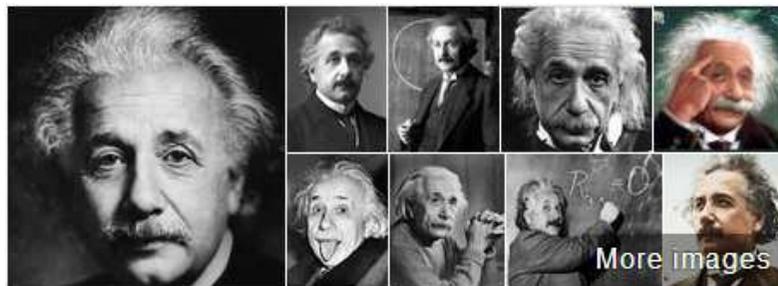
Bonn Berlin



- [Bahn von Bonn nach Berlin - VerkehrsmittelVergleich.de](#)
www.verkehrsmittelvergleich.de/bahn/bonn/berlin ▾
 Günstig mit der Bahn von Bonn nach Berlin und dabei alle Bahnverbindungen mit anderen Verehrsmitteln vergleichen.
- [Flug von Köln-Bonn nach Berlin-Tegel ab 33,00 - Germanwings](#)
www.germanwings.com > ... > [Flug nach Berlin-Tegel](#) ▾
 Günstiger Flug von Köln-Bonn (CGN) nach Berlin-Tegel (TXL) ab 33,00 €* - hier buchen.
 Mit Germanwings fliegen und bereits 72h vor Abflug einchecken.
- [Bonn-Berlin. Das Debattentagebuch im Internet](#)
www.bonn-berlin-debatte.de/blog/ ▾
 08.11.2012 – Das Blog befasst sich mit der Debatte über den Umzug der Bundesregierung von Bonn nach Berlin.
- [Flug Köln / Bonn Berlin \(CGN-BER\) - airberlin.com](#)
www.airberlin.com/de-DE/destinations/routes/flight/cgn-ber ▾
 Flug Köln/Bonn - Berlin (CGN-BER): die schnelle Verbindung in die Hauptstadt. Einen günstigen Flug von Köln/Bonn nach Berlin finden Sie bei airberlin.
- [Flug Köln-Bonn-Berlin Tegel - Fluege.de](#)
www.fluege.de > ... > [Flüge Berlin](#) > [Flughafen Berlin-Tegel](#) ▾
 Buchen Sie Ihren Flug von Köln-Bonn nach Berlin Tegel bei Fluege.de und vergleichen Sie die Flüge von Köln-Bonn nach Berlin Tegel von über 750 Airlines ...

Retrieved Snippet:

Google Knowledge Graph



Albert Einstein

Albert Einstein was a German-born theoretical physicist who developed the general theory of relativity, effecting a revolution in physics. [Wikipedia](#)

Born: March 14, 1879, Ulm

Died: April 18, 1955, Princeton

Children: Hans Albert Einstein, Eduard Einstein, Lieserl Einstein

Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)

Education: University of Zurich (1905), ETH Zurich (1901), [More](#)

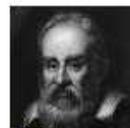
People also search for



Isaac Newton



Stephen Hawking



Galileo Galilei



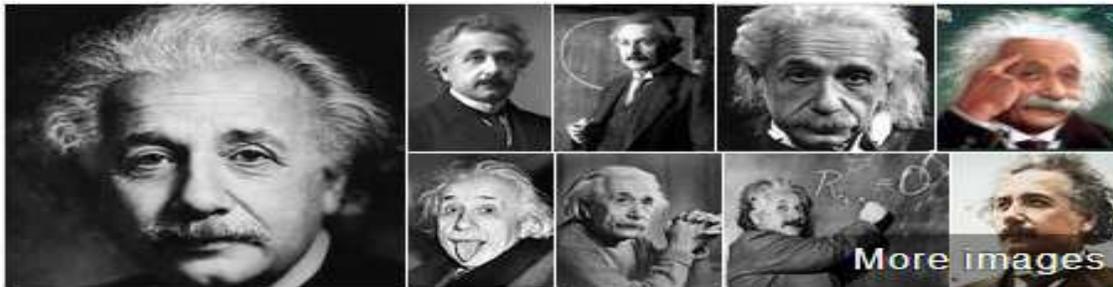
Thomas Edison



Niels Henrik David Bohr

Is there a “Why” Button?

“Lessons learned” from Google Knowledge Graph



Albert Einstein

Albert Einstein was a German-born theoretical physicist who developed the general theory of relativity, effecting a revolution in physics. [Wikipedia](#)

Born: March 14, 1879, Ulm

Died: April 18, 1955, Princeton

Children: Hans Albert Einstein, Eduard Einstein, Lieserl Einstein

Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)

Education: University of Zurich (1905), ETH Zurich (1901) [More](#)

People also search for



Isaac Newton



Stephen Hawking



Galileo Galilei



Thomas Edison



Niels Henrik David Bohr

What does Thomas Edison have to do with Einstein ?

User needs to explore by itself and there is no initial headwords to start with!

Challenging Tasks

- connecting two points of knowledge (entities)
- *describing the “connections”*
 - explaining the relationship
 - giving meaningful, human-readable information
 - unexpectedness, usefulness (serendipity)
- ...

Making sense of connections?

- Some terms / entities are more connected than others
 - “Crime” and “victim” are close
 - “Crime” and “love” not so much
- For text analysis we need to know how much they are related (on a scale from 0..1 preferably)
- For humans, we are also interested in what sense are they related (*in text preferably*)

“Semantic Relatedness”:

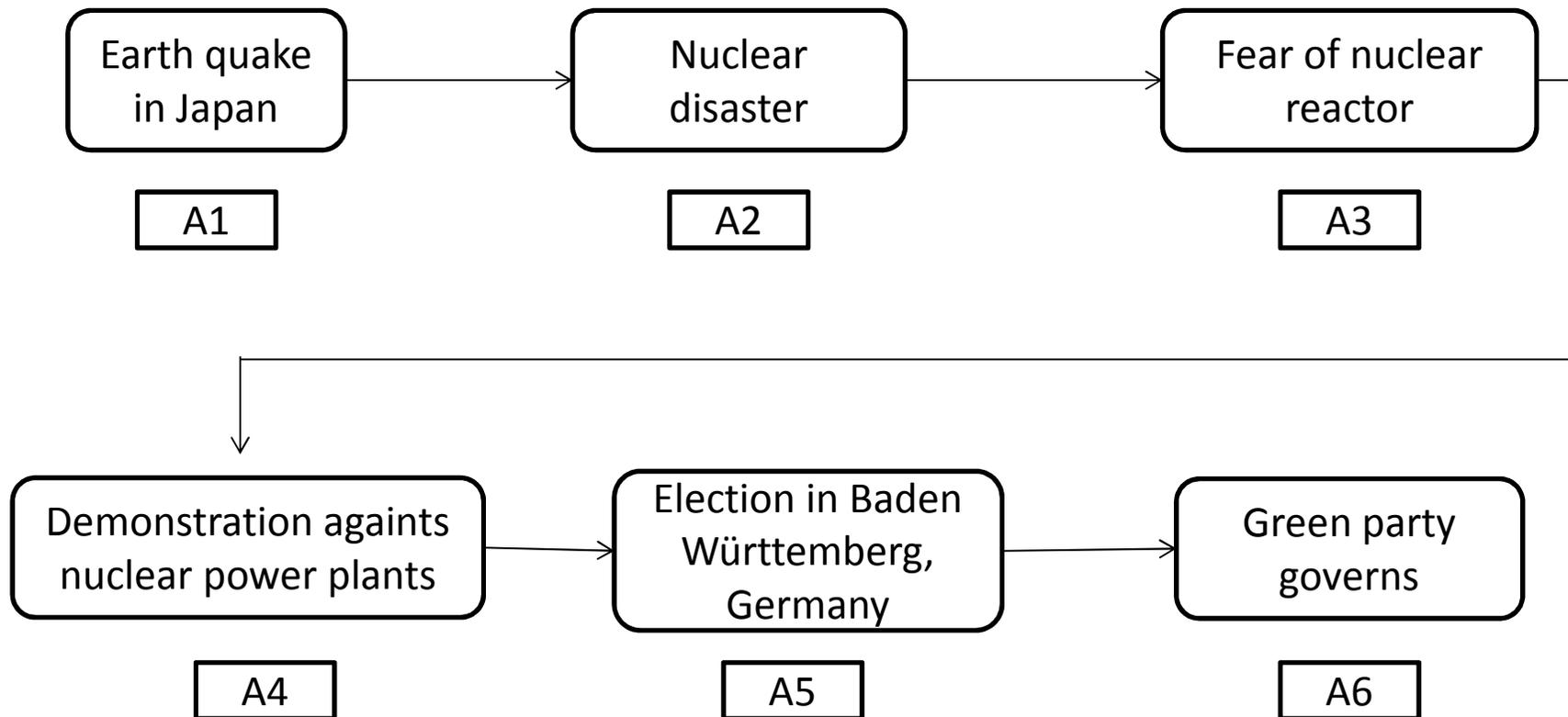
- connecting two entities:
 - A and B
 - Result: xy% similar

“Semantic Relatedness ++”

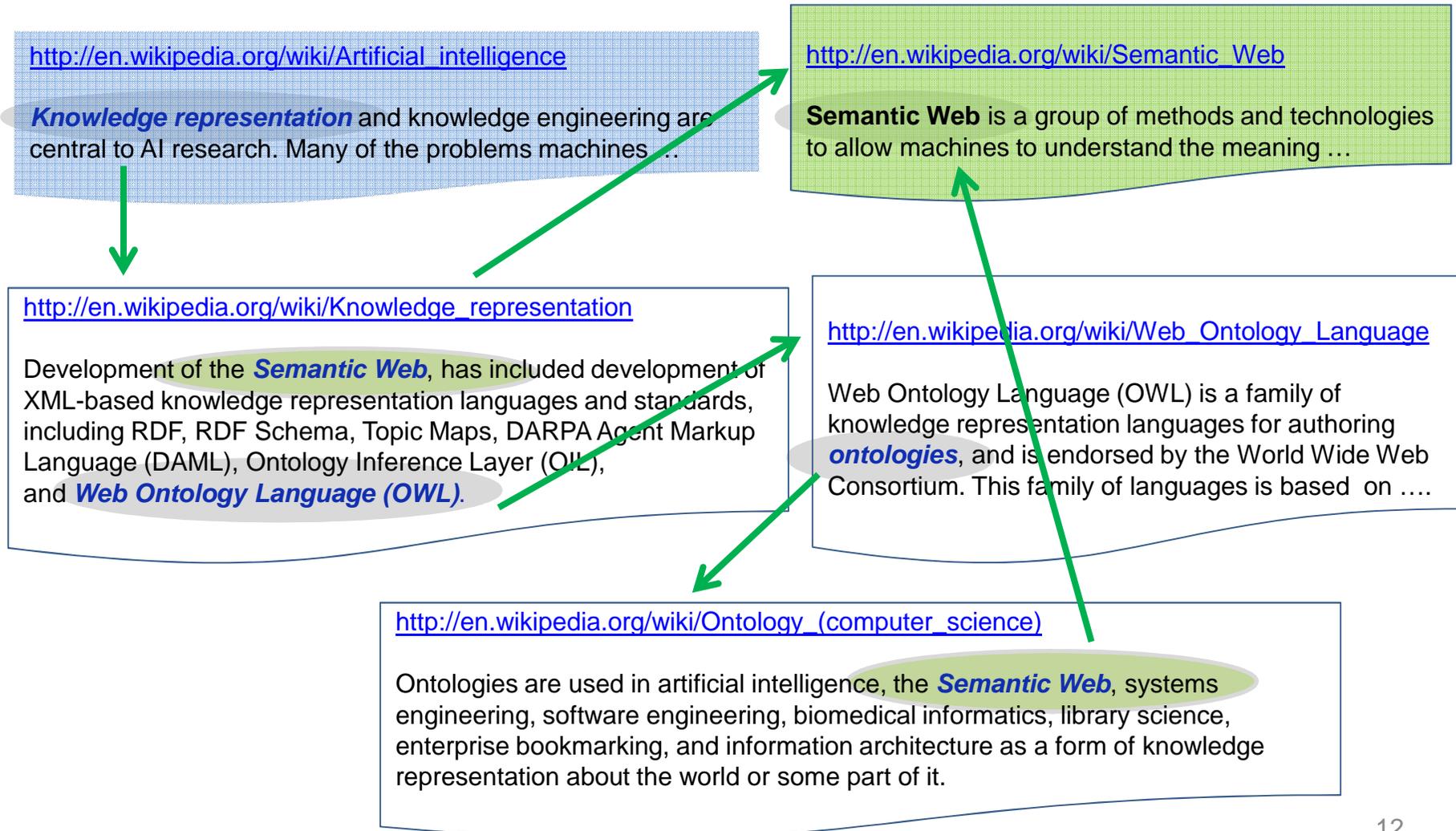
- connecting two entities:
 - A and B
 - Result: xy% similar
 - describing the “relatedness”
 - (some explaining texts)

- Relationship between two entities:
 - **Radio** and **Television**
 - Result: 70% related
- ← helpful?
- Relationship (+ explanation) between two entities:
 - **Radio** and **Television**
 - Result:
 - 70% related
 - *Television sends the picture as AM and the sound as AM or FM, with the sound carrier a fixed frequency (4.5 MHz in the NTSC system) away from the video carrier.*

Connecting the Dots: Chain of News Articles



Desired Results



Important References

- Extracting Knowledge / Wikipedia (Wira-Alam, Mathiak, Zloch 2012 – 2013)
- Connecting the Dots, Metro Maps, Large Scale Maps of Information (Shahaf et al., 2010 – 2013)
- Serendipitous Entity Search (Bordino et al., 2013)
- Improving Search Result Summaries (Ageev et al., 2013)

Wikipedia Datasets

Wikipedia (EN)

Page Articles	10,276,554
Page-to-Page Links	315,255,359

Wikipedia (DE)

Page Articles	2,673,528
Page-to-Page Links	82,515,479

Wikischools for Filtering (School related Wikipedia pages)

Short List	1,875
Complete List	8,944

Start : Albert Einstein

End : Thomas Edison

Distance : max 3 hops

The screenshot displays the 'gesis Delphin Search' interface. The main area shows a network graph with nodes representing concepts and individuals. The nodes are: Albert_Einstein (blue oval), Thermodynamics (yellow rounded rectangle), James_Watt (yellow rounded rectangle), Radio (yellow rounded rectangle), Electrical_engineering (yellow rounded rectangle), Alternating_current (yellow rounded rectangle), and Thomas_Edison (green oval). Edges connect Albert_Einstein to Thermodynamics, Electrical_engineering, and Alternating_current; Thermodynamics to James_Watt; James_Watt to Thomas_Edison; Radio to Thomas_Edison; Electrical_engineering to Radio and Alternating_current; and Alternating_current to Thomas_Edison.

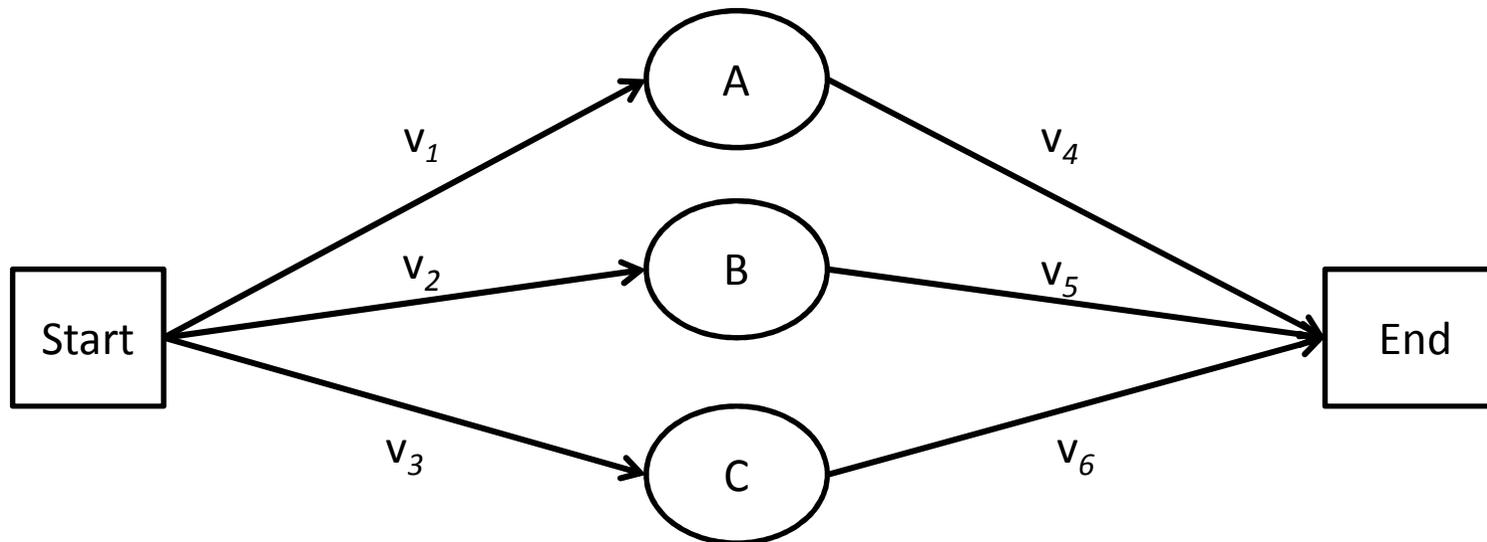
On the right side, the 'Routes API' control panel is visible. It includes input fields for 'Title 1: Albert Einstein' and 'Title 2: Thomas Edison', a 'Distance' field set to '3', and buttons for 'redraw' and 'visualize'. Below this, the 'Results' section shows a list of paths:

```

1. Albert_Einstein -> Alternating_current
2. Albert_Einstein -> Electrical_engineer
3. Albert_Einstein -> Electrical_engineer
4. Albert_Einstein -> Electrical_engineer
5. Albert_Einstein -> Thermodynamics -

```

start to end: How to reach the end?



Choosing a highest reachability score. This score describes a “probability” of reaching a node *end* given a node *start*.

Similarity Score between two nodes A and B:

$$sim(A, B) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

Reachability between two nodes A and B (in a path):

$$r_{ij} = \prod_i^{j-1} sim(term_i, term_{i+1}) \quad (2)$$

How to rank the excerpts:

- Based on position on the text, e.g. abstracts have “best score”

- Based on normalized term frequency
 - Excerpt: Berlin is the capital of Germany
 - Stopwords: is, the, of
 - Important words: Berlin, capital, Germany

$$score = \frac{freq_{Berlin} + freq_{capital} + freq_{Germany}}{\# words \setminus \# stopwords}$$

Short Demo

<http://tinyurl.com/knowescape-demo-v3>

Some Points....

- help users gain a better understanding about connections between point of knowledge
- while preliminary, suggests a promising direction for further investigation
- challenge for us to improve the current results

Thank You for Your Attention!

*I am looking forward to your remarks,
questions, critics, feedback.*