

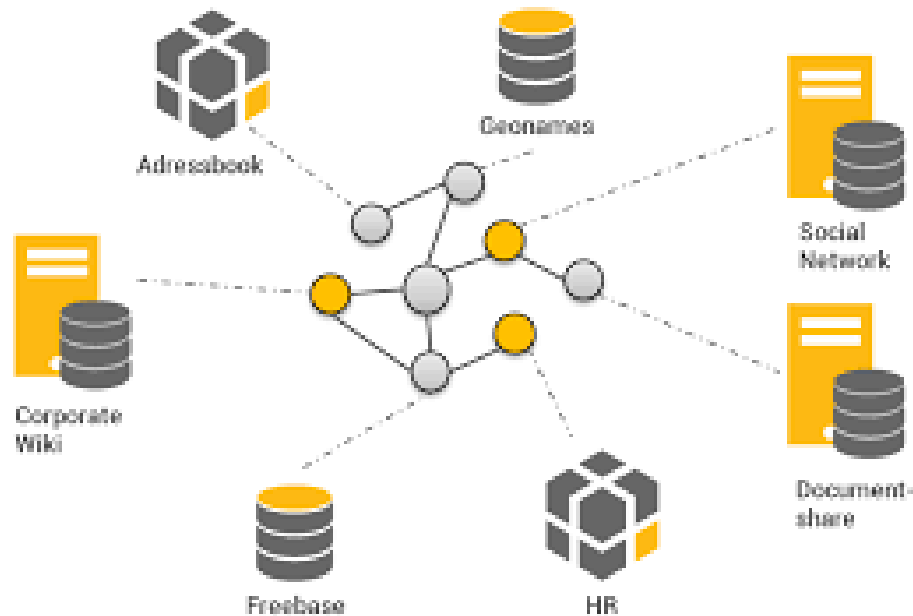
Knowledge Graphs

Seminar Kick Off



Goals and Motivation

- Knowledge Graphs are used...
- ...in companies and organizations
 - collect and organize knowledge in the company
 - link isolated information sources
 - make information searchable and findable



Masuch, 2014

Goals and Motivation

- Knowledge Graphs are used...
- ...as (free), public resources
 - collect common knowledge
 - general purpose, not task specific
 - make it easy to build knowledge-intensive applications



Goals and Motivation

- The one knowledge graph you probably know:
 - Google's knowledge graph
 - Google also coined the term

The screenshot shows a Google search for "nine inch nails". At the top, there's a search bar with the text "nine inch nails" and a magnifying glass icon. Below the search bar, there are tabs for "All", "Videos", "Images", "News", "Shopping", "More", "Settings", and "Tools". The search results show "About 23,600,000 results (0.52 seconds)".

Below the search results, there's a privacy reminder from Google with a shield icon and a "REVIEW" button. Below that, there are links to the "nine inch nails | the official website" and "Nine Inch Nails - Wikipedia".

The Wikipedia link leads to a page that says "Nine Inch Nails, commonly abbreviated as NIN (stylized as NIN!), is an American industrial rock band from Cleveland, Ohio, founded in 1988. The band consists ...". It also lists "Years active: 1988–present", "Origin: Cleveland, Ohio, U.S", "Members: Trent Reznor, Atticus Ross", and "Labels: Nothing, TVT, Interscope, The Null Cor...".

Below the Wikipedia link, there's a "Videos" section with three video thumbnails. The first video is titled "Nine Inch Nails – LESS THAN" and has a duration of 3:33. The second video is titled "Nine Inch Nails – The Hand That Feeds" and has a duration of 3:43. The third video is titled "Nine Inch Nails – Head Like A Hole" and has a duration of 4:32.

On the right side of the search results, there's a knowledge panel for "Nine Inch Nails". It features a collage of images of the band members and the band's logo. Below the images, it says "Nine Inch Nails" and "Rock band". There's a link to "nin.com".

Below the link to "nin.com", there's a section titled "Available on" with logos for YouTube, Spotify, and Deezer. Below this, there's a paragraph of text about the band: "Nine Inch Nails, commonly abbreviated as NIN, is an American industrial rock band from Cleveland, Ohio, founded in 1988. The band consists of producer and multi-instrumentalist Trent Reznor, as well as English musician Atticus Ross. [Wikipedia](#)".

Below the paragraph, there's a section titled "Origin: Cleveland, Ohio, United States (1988)". Below that, there's a section titled "Members: Trent Reznor, Atticus Ross, Eric Avery, Robin Finck, MORE". Below this, there's a section titled "Genres: Industrial rock, Industrial music, Alternative rock, Electronic rock". Below that, there's a section titled "Record labels: Interscope Records, Universal Music Group, MORE".

Organization

- Requirements
 - Familiarize yourself with a particular knowledge graph and present it in the seminar
 - Write a seminar paper
 - Review others' seminar papers
 - it is a good idea to also read the main papers for the topics you review
- First step
 - Pick a knowledge graph
 - If not done yet, send a ranked list to Ms. Bianca Lerner

Organization

- We will use a process called “peer review”
 - widely used (and discussed) in science
 - you will review your fellow students’ seminar papers
- Timeline
 - Prepare a draft until April 7th
 - You will get two seminar papers to review
 - Submit your reviews until April 21th
- Seminar (i.e. ,presentations, discussions)
 - April 29th, May 6th/13th/20th, 13.45-17.00, room A3.01
 - Participate *actively*
- Final seminar paper submission: June 30th

Preparing Your Seminar Paper

- Read the original paper
 - get the idea, read works on relevant foundations
- Read follow up papers about the topic
 - what newer developments have there been?
- Find papers citing the original paper
 - who says what about it?
 - who uses it, and for which purposes?
- Anything else you want to discuss
- Key takeaway:
 - you are not done after reading the original paper
 - use further literature!

Preparing a Review

- 1st rule: be constructive!
- What you should point at
 - can you follow easily? is there information missing at any point?
 - are all claims well supported?
 - do you have any questions not answered?
 - aspects underrepresented
- What you should not do
 - provide general criticism (“don’t like the paper”)
 - correct every spelling mistake
 - rewrite the seminar paper

Preparing the Presentation

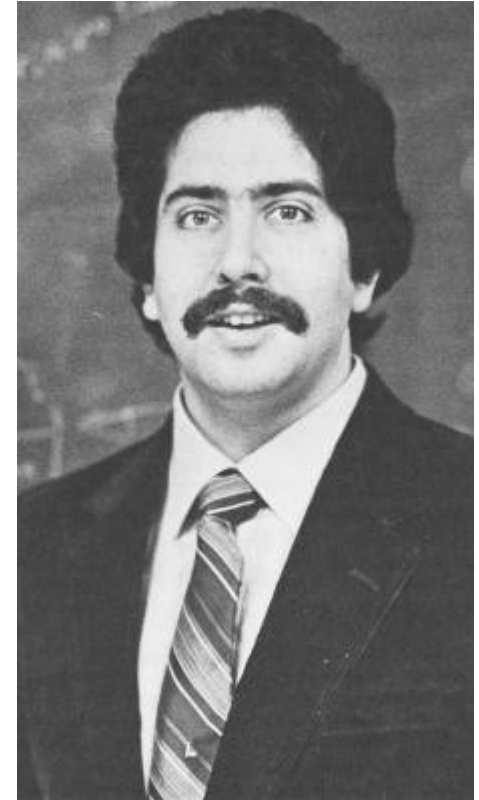
- You don't have to start at zero
 - your fellow students are also familiar with knowledge graphs
- Focus on key aspects
 - what is special about the graph you present?
 - what are the novel ideas?
- Be illustrative
 - use examples
 - show examples of actual data, if available
- Be entertaining
 - it's (partly) up to you whether we are having fun here ;-)

Topic Overview

- Sources for general purpose knowledge graphs:
 - Manual curation
 - Semi-structured knowledge, such as Wikipedia
 - Unstructured text or web page collections
 - ...

Manual Curation

- Cyc
- Encyclopedic collection of knowledge
 - Started by Douglas Lenat in 1984
 - Estimation: 350 person years and 250,000 rules should do the job of collecting the essence of the world's knowledge
- Today
 - a knowledge graph with 21M facts, various modules
 - estimated development cost: 120M\$



Manual Curation

- Freebase
 - Collaborative editing
 - Schema not fixed
- Today
 - Acquired by Google in 2010
 - Powered first version of Google's Knowledge Graph
 - Shut down in 2016
 - Partly lives on in Wikidata (see in a minute)



Manual Curation

- Wikidata
 - launched 2012
 - Goal: centralize data from Wikipedia languages
 - Collaborative
 - Imports other datasets
- Present
 - One of the largest public knowledge graphs
 - Includes rich provenance information



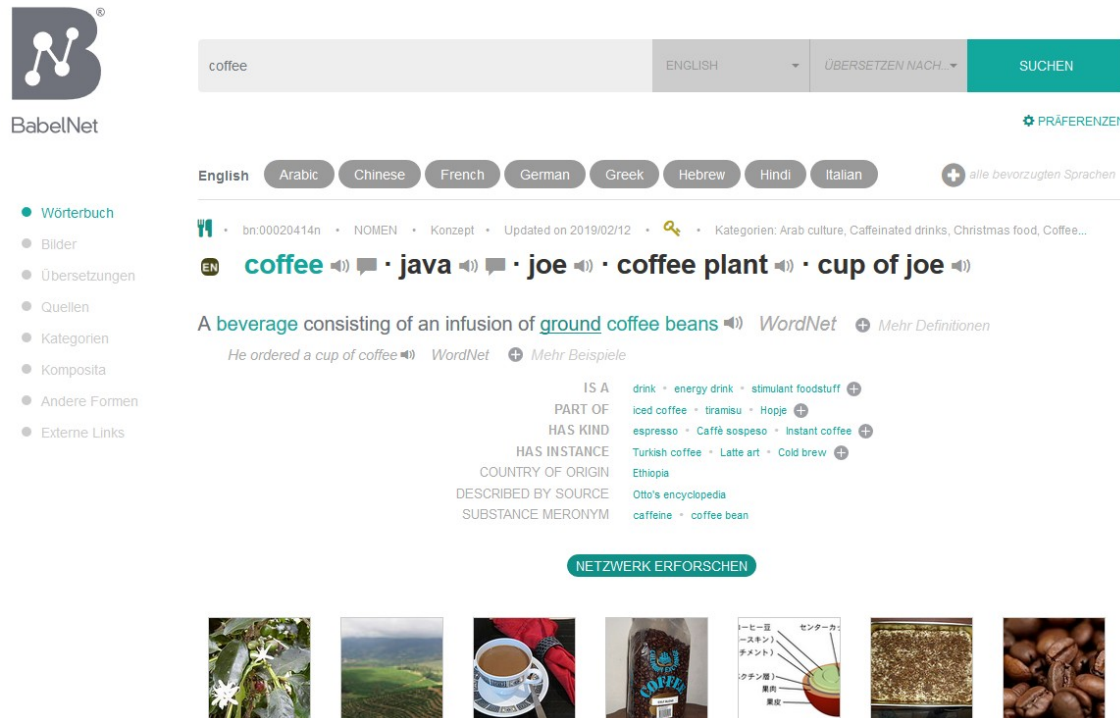
Creation from Wikipedia

- The 2010s
 - DBpedia: launched 2007
 - YAGO: launched 2008
- Harvesting structured parts of Wikipedia, e.g.,
 - infoboxes
 - categories
 - geographic coordinates
- Present
 - two of the most used knowledge graphs



Creation from Wikipedia

- BabelNet
 - started in 2012
 - collects definitions in various languages
 - also integrates other linguistic sources and dictionaries



The screenshot displays the BabelNet interface for the word 'coffee'. At the top, there is a search bar with 'coffee' entered, a language dropdown set to 'ENGLISH', and a 'SUCHEN' button. Below the search bar, a row of language tabs includes English, Arabic, Chinese, French, German, Greek, Hebrew, Hindi, and Italian, along with a link to 'alle bevorzugten Sprachen'. The left sidebar contains a 'Wörterbuch' section with links to 'Bilder', 'Übersetzungen', 'Quellen', 'Kategorien', 'Komposita', 'Andere Formen', and 'Externe Links'. The main content area shows the word 'coffee' with its various forms and translations: 'java', 'joe', 'coffee plant', and 'cup of joe'. It also provides a definition: 'A beverage consisting of an infusion of ground coffee beans'. Below this, there are several semantic relations: 'IS A' (drink, energy drink, stimulant foodstuff), 'PART OF' (iced coffee, tiramisu, Hopje), 'HAS KIND' (espresso, Caffè sospeso, Instant coffee), 'HAS INSTANCE' (Turkish coffee, Latte art, Cold brew), 'COUNTRY OF ORIGIN' (Ethiopia), 'DESCRIBED BY SOURCE' (Otto's encyclopedia), and 'SUBSTANCE MERONYM' (caffeine, coffee bean). A 'NETZWERK ERFORSCHEN' button is located below the relations. At the bottom, there are seven small images related to coffee: coffee plants, a coffee landscape, a cup of coffee, a bag of coffee, a coffee machine, coffee beans, and a close-up of coffee beans.

Creation from Wikipedia

- ConceptNet
 - uses multiple sources (e.g., Wiktionary, DBpedia, dictionaries)
 - multiple languages
 - linguistic frames (related to, caused by, requires ...)
 - used for training word sense embeddings (similar to word2vec)



An English term in ConceptNet 5.6

Sources: Open Mind Common Sense contributors, DBpedia 2015, JMDict 1.07, Verbosity players, German Wiktionary, English Wiktionary, French Wiktionary, and Open Multilingual WordNet
View this term in the API

[Documentation](#) [FAQ](#) [Chat](#) [Blog](#)

Synonyms	Related terms	Causes of knowledge	Things that require knowledge
sh znanje →	en learn →	en learning →	en answering questions →
ja ナレッジ (n) →	en acknowledge →	en remembering →	en checking vital signs →
en cognition (n) →	en driver →	en listening to the radio →	en testing the car →
ja ノリッジ (n) →	en confess →	en memorising →	en weeding the garden →
af kennis →	en london →	en study →	en debating politics →
ja ノレッジ (n) →	sh znanje (n) →	en studying →	en designing software →
ar علم →	en science →	en analysing something →	en discovering the truth →
ja 人智 (n) →	en data →	en discovering the truth →	en doing a crossword puzzle →
ja 人知 (n) →	en route →	en expressing information →	en expressing information →
ar معرفة →	en erudition →	en finding information →	en fix a computer →
ast conocencia →	en test →	en going to school →	en giving a clue →
ja 学 (n) →	de archaisch →	en knowing how the stock market performed →	en handling proposals →
ia 学力 (n) →	en information →		en having an examination →

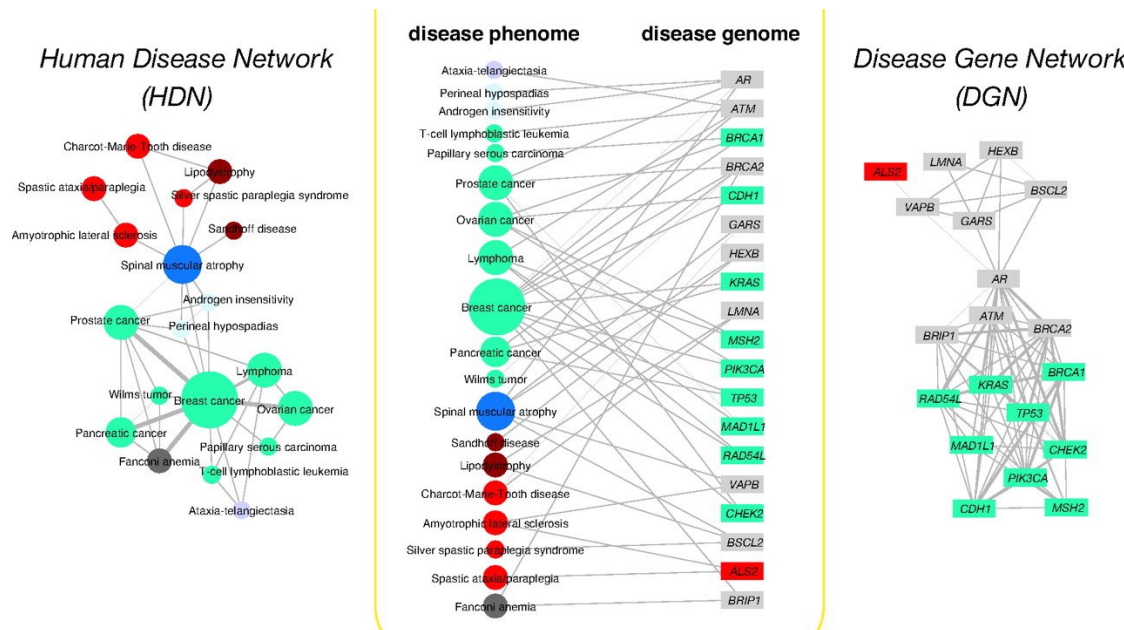
Creation from Unstructured Text / Web Sources

- NELL
 - a system running at CMU ever since January 2010
 - started with a few facts and patterns
 - learns to extract facts from the Web by
 - ...applying the patterns it knows
 - ...generalizing facts to new patterns
 - new updates provided every other day



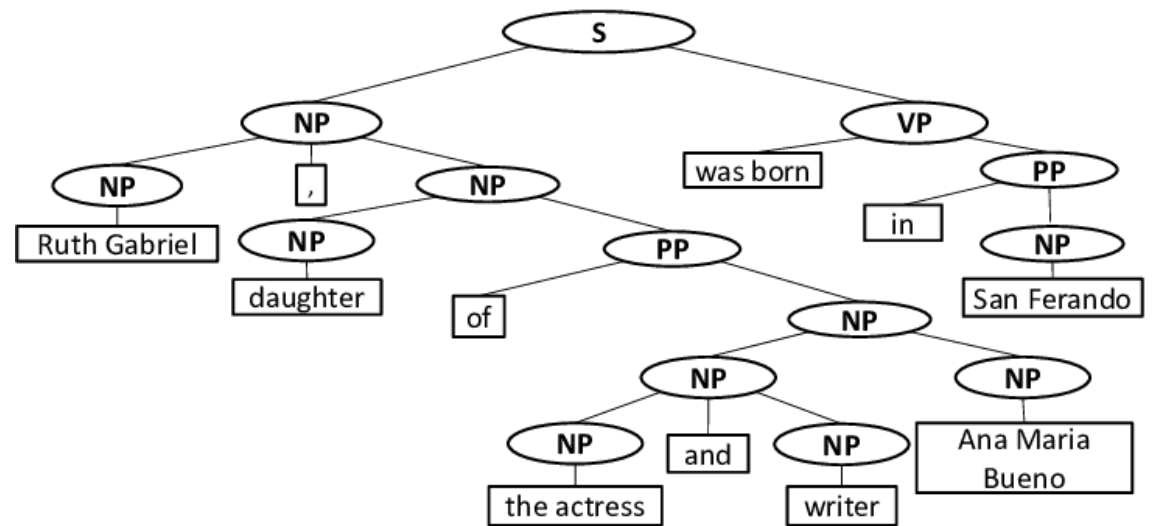
Creation from Unstructured Text / Web Sources

- DeepDive
 - created at Stanford University
 - technology for learning facts from unstructured sources
 - combines NLP, machine learning, logical inference
 - used in various domains



Creation from Unstructured Text / Web Sources

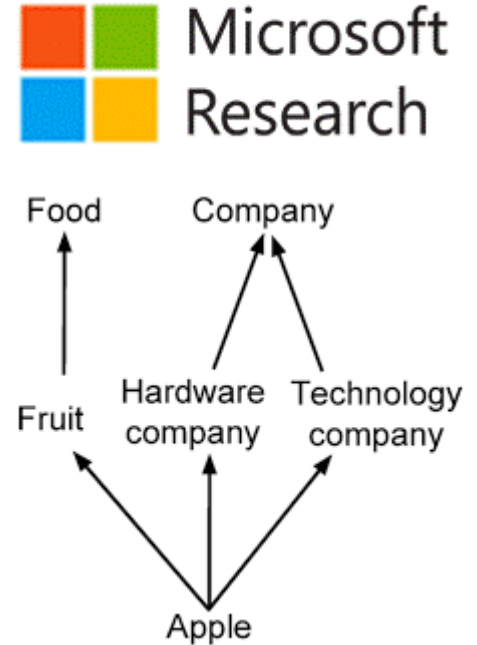
- ReVerb / KnowItAll
 - NELL works with *closed* information extraction
 - ReVerb: *open* information extraction
 - No fixed schema
 - More flexible, but also more challenging



Bast, 2013

Creation from Unstructured Text / Web Sources

- Probase
 - a taxonomic knowledge graph
 - built by Microsoft Research
 - used in Bing Web search
 - learned from a huge Web corpus
- WebIsA(LOD)
 - open variant of Probase
 - built at DWS group, University of Mannheim
- Both contain
 - millions of facts
 - probabilities or trust scores



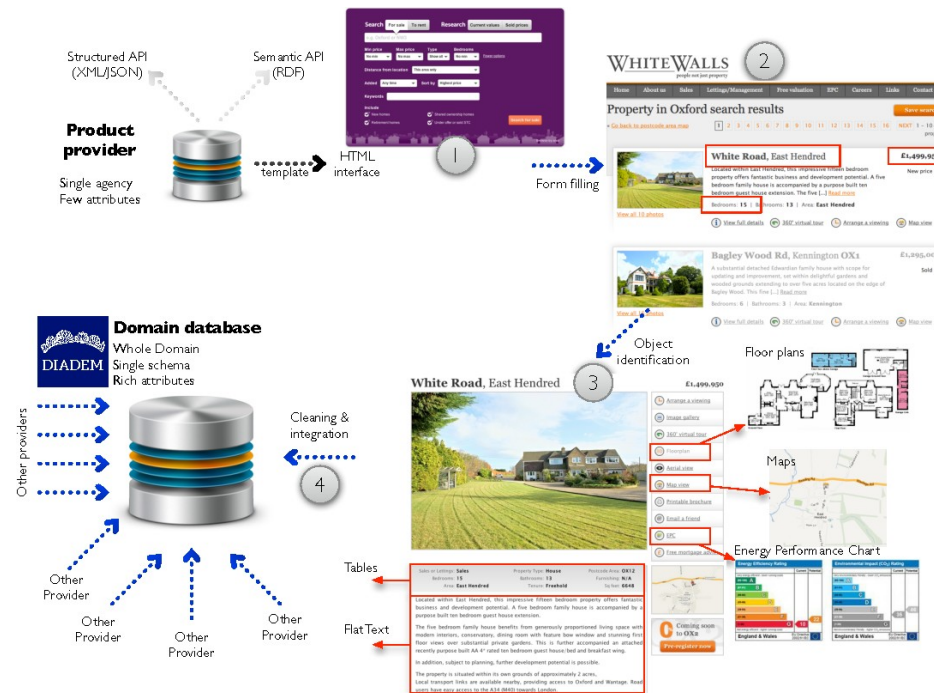
Creation from Unstructured Text / Web Sources

- KnowledgeVault
 - created at Google
 - next generation of Google's KG
 - various sources:
 - text and HTML sources
 - HTML tables
 - annotations (schema.org etc.)



Creation from Unstructured Text / Web Sources

- DIADEM
 - explores the DeepWeb
 - i.e., non-crawlable Web behind interfaces
 - reverse engineering knowledge bases



Furche, 2012

The Choice is Yours...

- compile a ranked list (at least three topics)
- send to Bianca Lerner by Wednesday eob
 - email: lermer@informatik.uni-mannheim.de
- you'll be assigned a topic by the end of the week



Questions?

