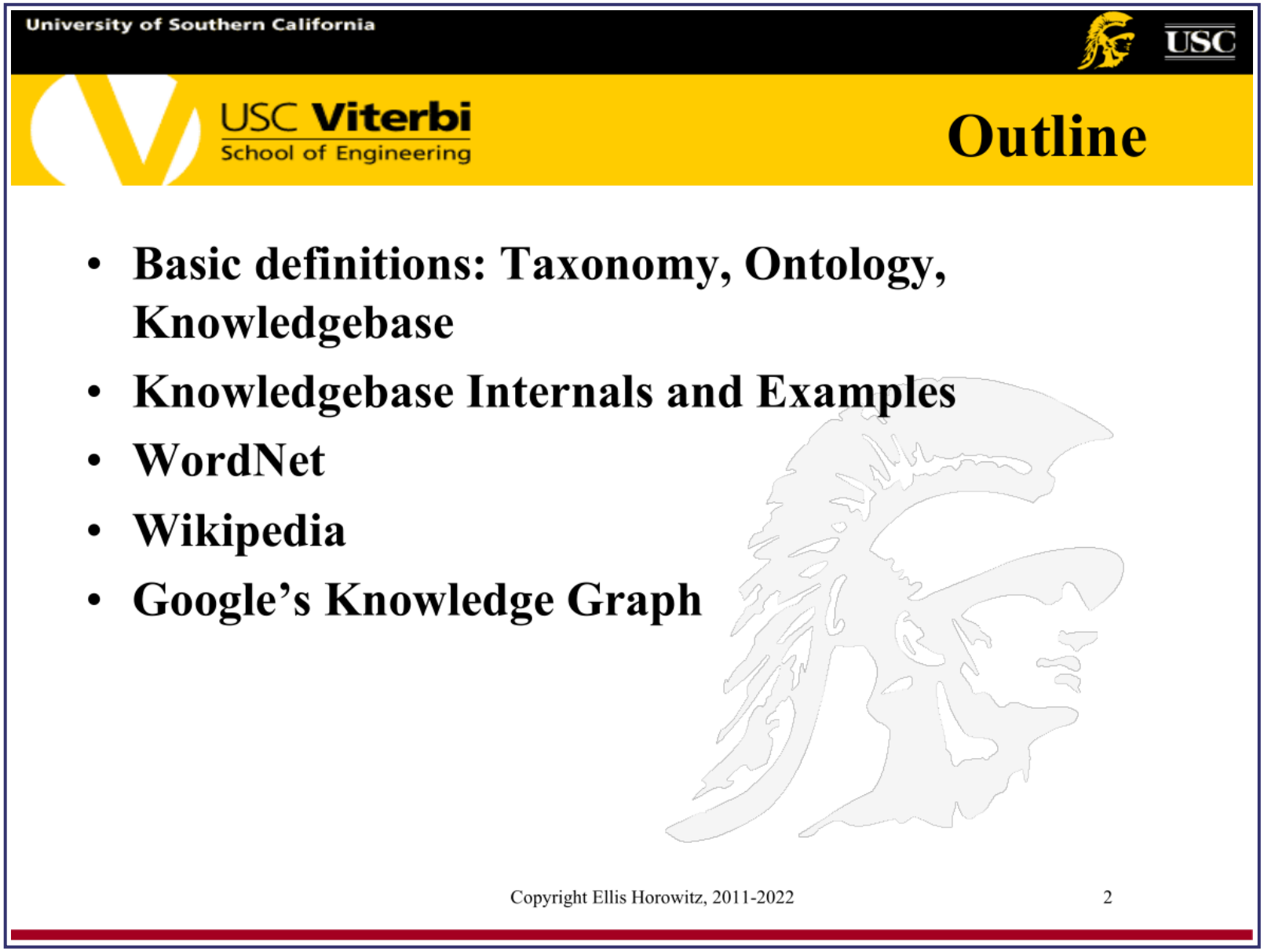
Knowledge graphs

USC's work: https://usc-isi-i2.github.io/home

••


University of Southern California
USC
USC Viterbi
School of Engineering
Outline
• Basic definitions: Taxonomy, Ontology, Knowledgebase
• Knowledgebase Internals and Examples
• WordNet
• Wikipedia
• Google's Knowledge Graph
Copyright Ellis Horowitz, 2011-2022
2

••

University of Southern California

USC Viterbi School of Engineering

What is a Taxonomy

What is a taxonomy

ταξις, ***taxis***
"arrangement"

+

νομος, ***nomos***
"law"

A taxonomy is a classification or categorization of a complex system.

e.g. **the ACM Computing Classification System, https://dl.acm.org/ccs**
the Mathematics Subject Classification

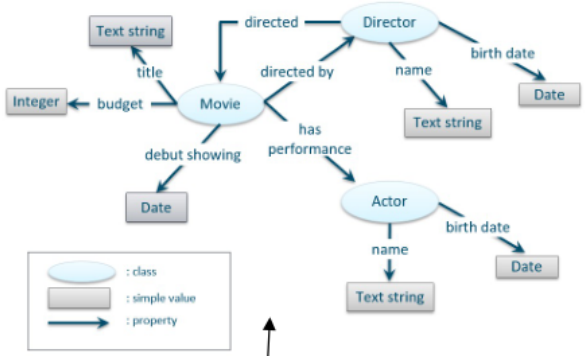
●●

University of Southern California

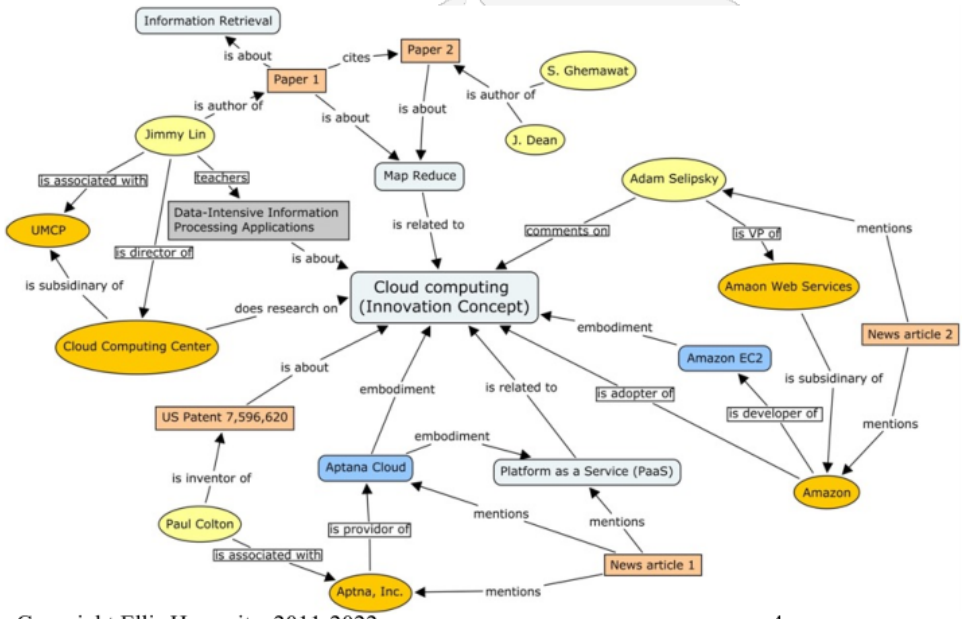
USC Viterbi School of Engineering

What is an Ontology

- a set of concepts and categories in a subject area or domain that shows their properties and the relations between them, or
- a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents, or
- a body of formally represented knowledge based on a *conceptualization*: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them



Brief ontology of movies



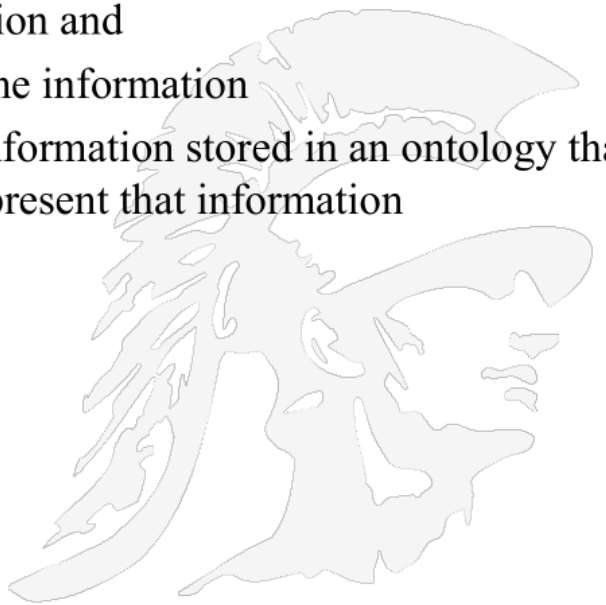
••

University of Southern California

USC Viterbi School of Engineering

What is a Knowledgebase

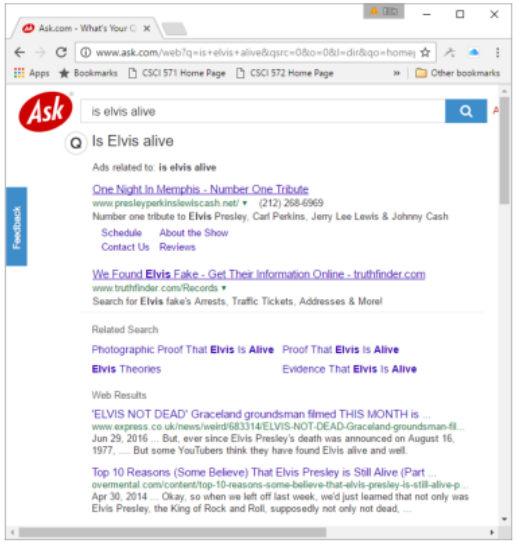
- A **knowledgebase** (**KB**) is a technology used to store and retrieve complex structured and unstructured information as stored in an ontology
  - Two components:
  - 1. a way of ***representing*** information and
  - 2. a method for ***reasoning*** about the information
- A **knowledgebase** is a collection of information stored in an ontology that includes software used to author and present that information



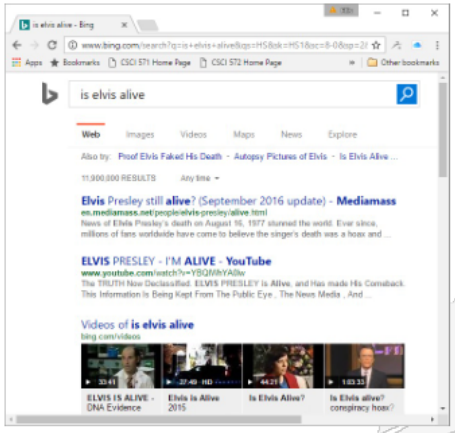
University of Southern California

USC Viterbi School of Engineering

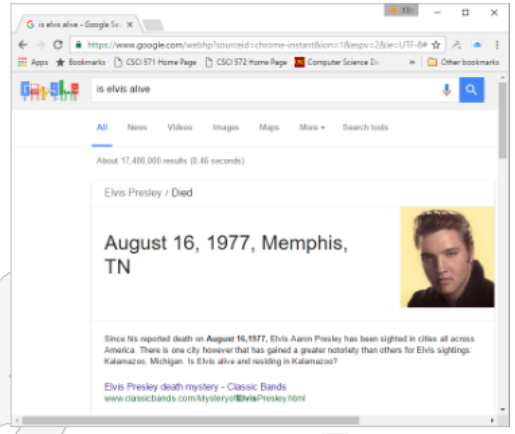
Is Elvis Alive



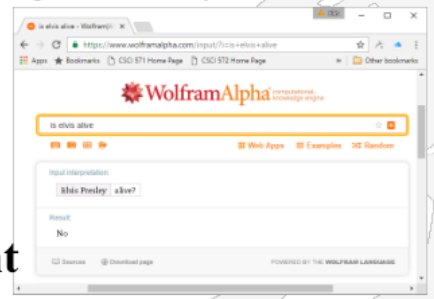
Ask.com – Mostly Yes



Bing.com – Yes/Maybe



Google.com - No



Wolfram Alpha - No

2 Yes'es and 2 No's;

text matching alone is insufficient

we need to "understand" the query

••

University of Southern California

USC Viterbi School of Engineering

KnowledgeBases-Developed by AI Community

- **To move away from just using keyword matching, search engines borrowed techniques developed by AI researchers**
- A ***knowledgebase*** (KB) is a technology used to store complex structured and unstructured information used by a computer system.
- **A knowledge-based system consists of two elements:**
  1. a *knowledgebase* that ***represents*** facts about the world, and
  2. an *inference engine* that can ***reason*** about those facts (and use rules and other forms of logic to deduce new facts or highlight inconsistencies)
- **The term "knowledgebase" was coined to distinguish this form of knowledge store from the more common and widely used term *database***
  - Is a relational database a knowledgebase?
- ***Examples of knowledgebases: Freebase, Google's Knowledge Graph, Apple's Siri, IBM's Watson***

••

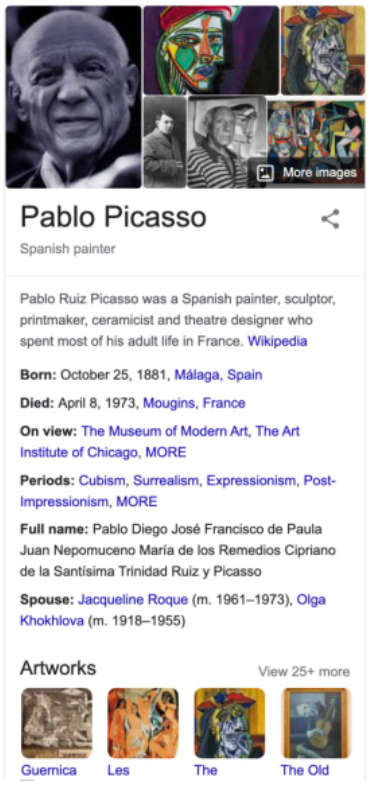
University of Southern California

USC Viterbi School of Engineering

USC

Search Engines Use Knowledgebases to Enhance the Display of Results

- **The representation of knowledge in a knowledgebase is an *object model***
  - Includes classes, subclasses and instances
- A **taxonomy** is usually only a hierarchy of concepts (i.e. the *only relation* between the concepts is parent/child, or subClass/superClass, or broader/narrower)
- In a **knowledgebase**, *arbitrary complex relations* between concepts can be expressed as well, e.g. (*X marriedTo Y*; or *A worksFor B*; or *C locatedIn D*, etc)
- **Search engines utilize this linked, structured data in many ways, such as**
  - Providing direct answers to queries
  - enhanced displays in many varieties of engaging visual formats, e.g. see query "Picasso" in Google



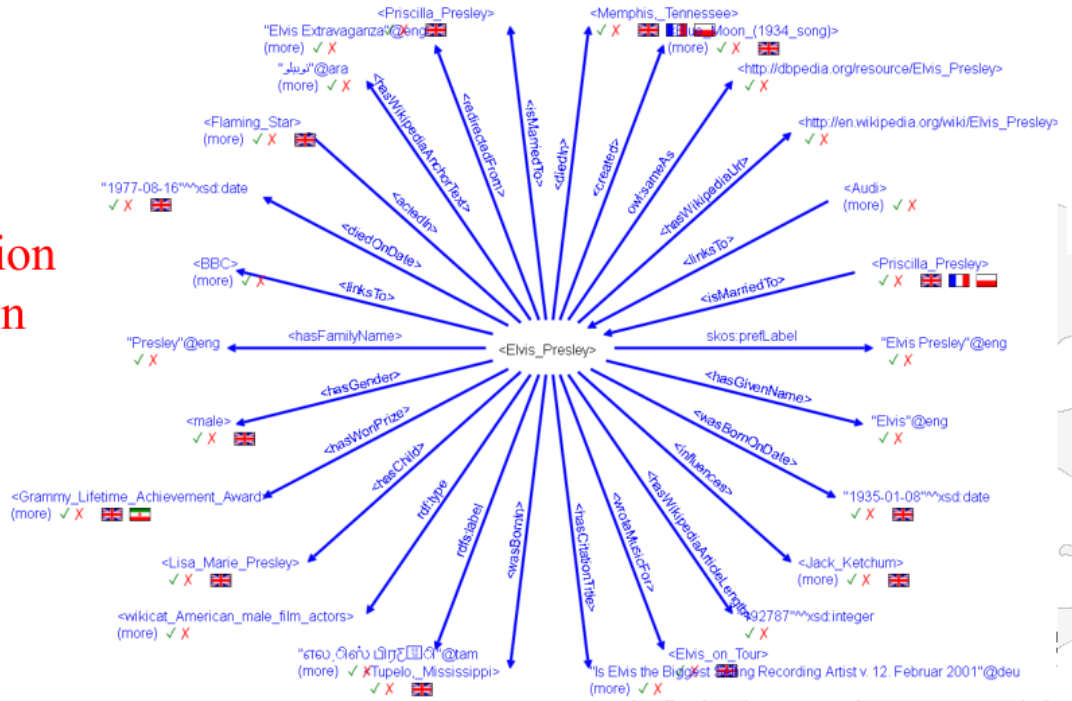
●●

University of Southern California

USC Viterbi School of Engineering

Close Up: A KnowledgeBase for Elvis

Information Extraction



Content analytics

"Elvis Presley, the first and greatest American rock-and roll star, died yesterday at the age of 42."

●●

University of Southern California

USC Viterbi School of Engineering

Types of Knowledge For a KnowledgeBase

Elvis Presley type American singer Elvis Presley type Baritone American singer subclassOf singer	taxonomic knowledge
Elvis Presley sang All Shook Up	factual knowledge
Elvis Presley bornIn Tupelo	
id11: Elvis Presley marriedTo Priscilla Presley id11 validDuring [1967, 1977]	temporal knowledge
Elvis Presley „has twin brother“ Jesse Garon	emerging knowledge
Elvis Presley „possibly has origin“ Cherokee	
Elvis Presley knownAs „The King of R&R“	terminological knowledge

Taxonomies are narrower than ontologies since ontologies include a larger variety of relation types. Mathematically, a hierarchical **taxonomy** is a tree structure of classifications for a given set of objects An ontology is a directed, labeled, cyclic graph.

••

University of Southern California

USC Viterbi School of Engineering

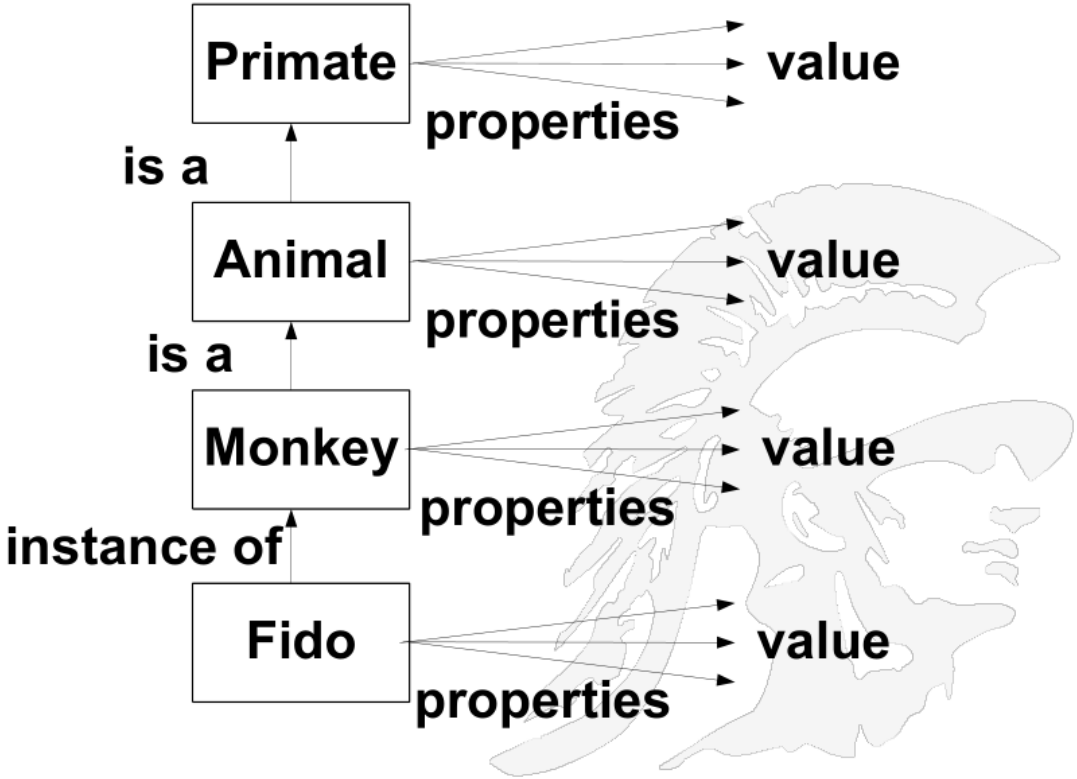
Inheritance

- An important feature of knowledge representation is its organization into *class hierarchies*. Classes can be based on the properties of objects/concepts
- Inheritance is a relation by which
  - 1. an individual assumes the properties of its class and
  - 2. determines which properties of a class are passed on to its subclass
- In order to support inheritance, objects/concepts must be organized into classes and classes must be arranged in a generalized hierarchy
  - the most generic object/concept is at the top of the hierarchy and the most specific is at the bottom of the hierarchy

University of Southern California

USC Viterbi School of Engineering

Inheritance

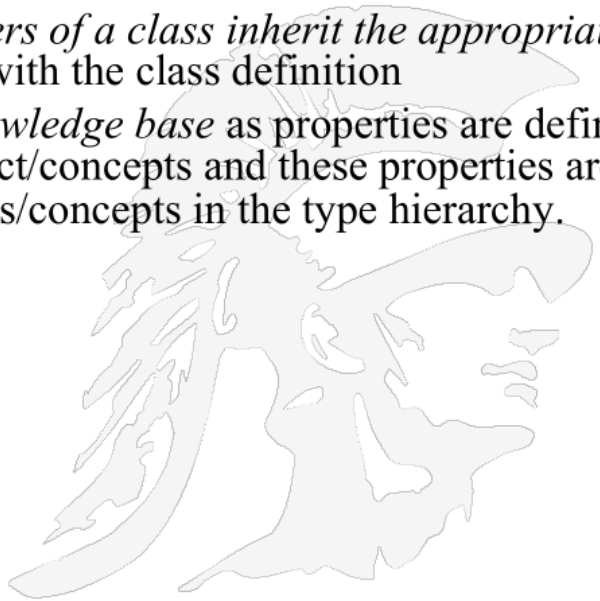


University of Southern California

USC Viterbi School of Engineering

Some Advantages of Inheritance

- Inheritance provides a natural mechanism for *representing taxonomically structured knowledge*
- Inheritance provides an economical means of *expressing properties common to a class of objects/concepts*
- Inheritance *guarantees that all members of a class inherit the appropriate properties* thus ensuring consistency with the class definition
- Inheritance *reduces the size of the knowledge base* as properties are defined once for the most general type of object/concepts and these properties are then share by other less generic objects/concepts in the type hierarchy.



●●

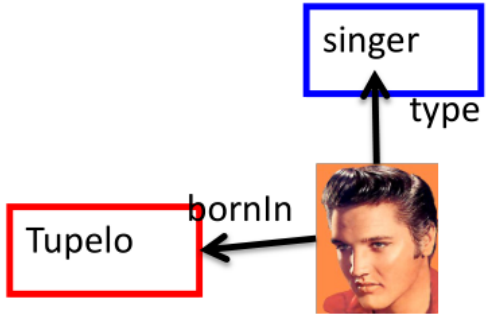
University of Southern California

USC Viterbi School of Engineering

Different Notations for a KnowledgeBase

- Resource Description Format (RDF) is a W3C spec used for creating ontologies;
  - https://www.w3.org/RDF/
  - Sometimes "RDF Ontology" and "KnowledgeBase (KB)" are used synonymously.

Graph notation:



Triple notation:

Subject	Predicate	Object
Elvis	type	singer
Elvis	bornIn	Tupelo
...	...	...

Logical notation:

type(Elvis, singer)
bornIn(Elvis,Tupelo)
...

●●

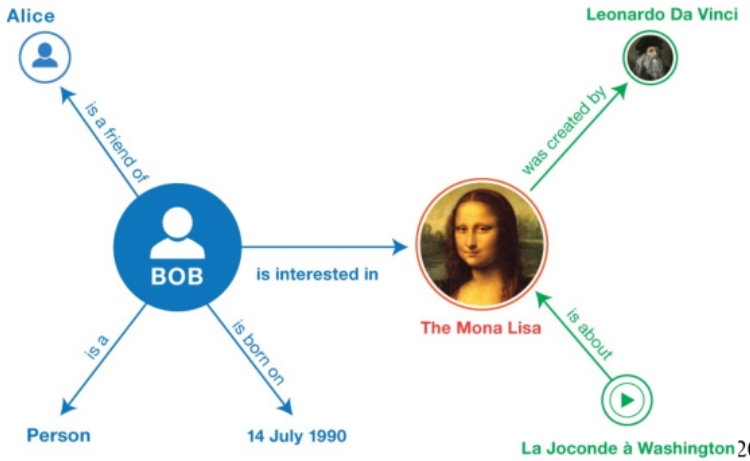
University of Southern California

USC Viterbi School of Engineering

RDF Data Model

- RDF allows us to make statements about resources.
  The format of these statements is:
  <subject> <predicate> <object>
- Some examples

<Bob> <is a> <person>
<Bob> <is a friend of> <Alice>
<Bob> <is born on> <the 4th of July 1990>
<Bob> <is interested in> <the Mona Lisa>
<the Mona Lisa> <was created by><Leonardo da Vinci>
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>



We can visualize triples as a connected **graph**.
Graphs consists of nodes and arcs.
The subjects and objects of the triples make up the nodes in the graph;
the predicates form the arcs

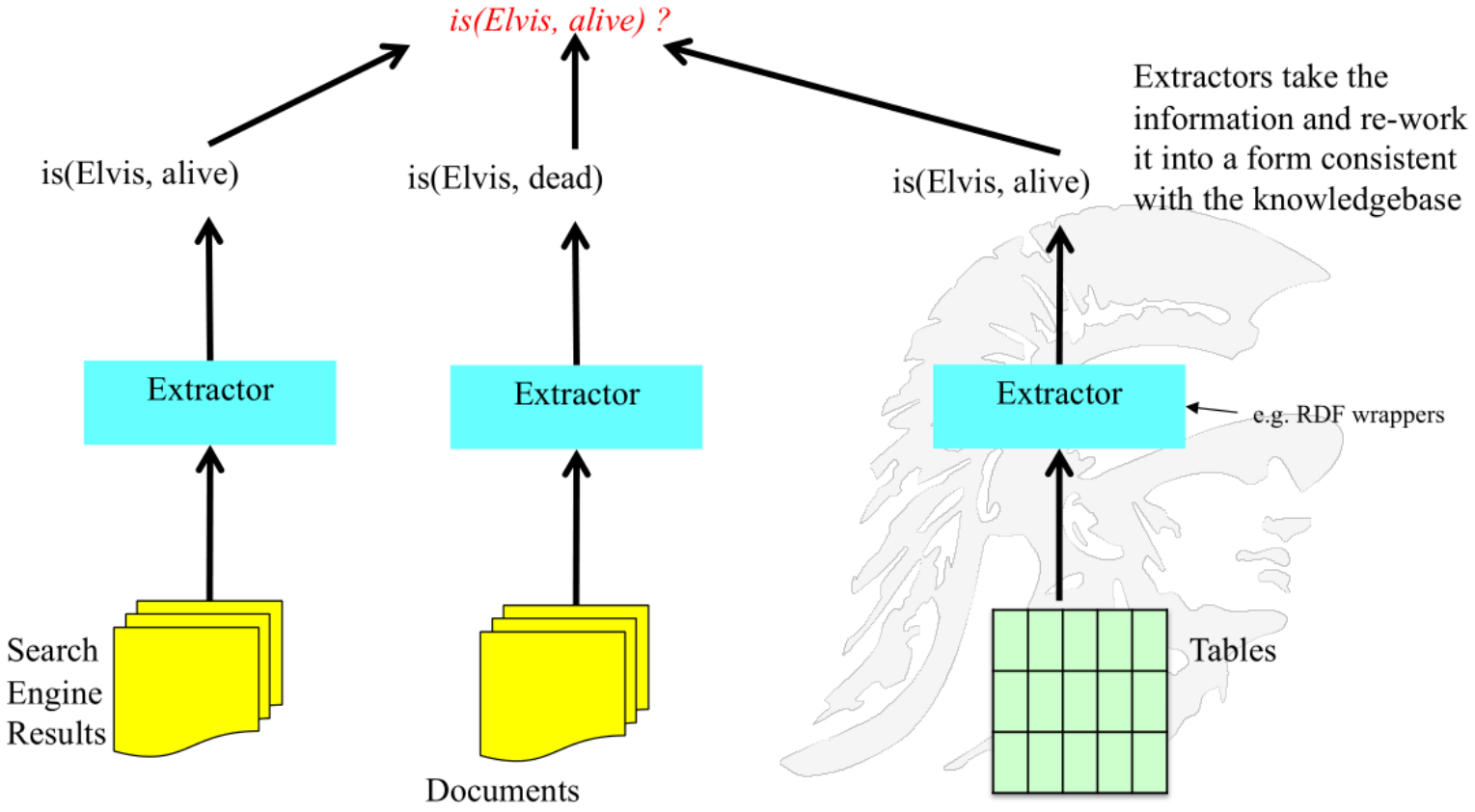
One query language for making inferences on these graphs is SPARQL

2011-2022

University of Southern California

USC Viterbi School of Engineering

To Answer a Question Knowledgebases Need to Combine Information From *Multiple Sources*

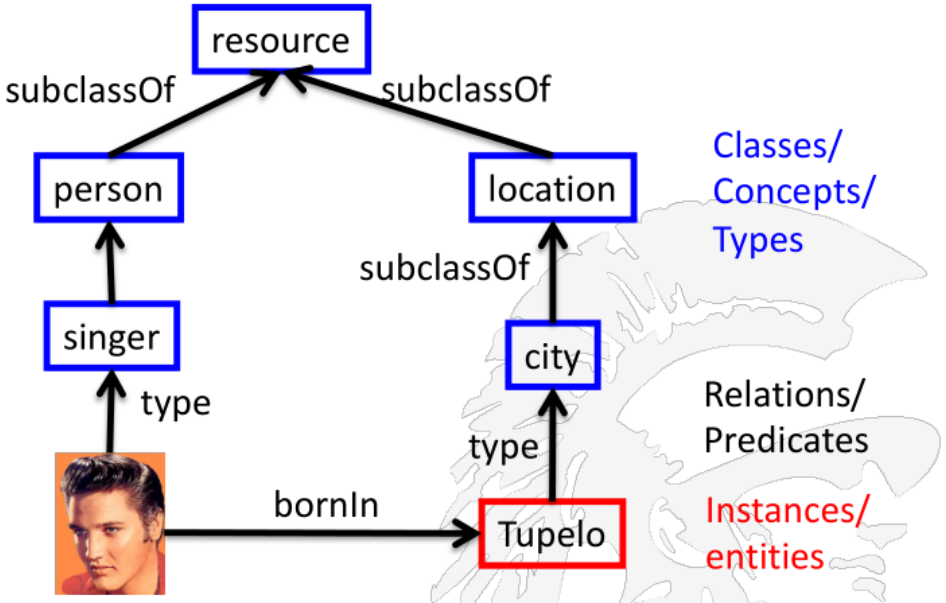


●●

University of Southern California

USC Viterbi School of Engineering

KnowledgeBases are Can Be Represented as Labeled MultiGraphs



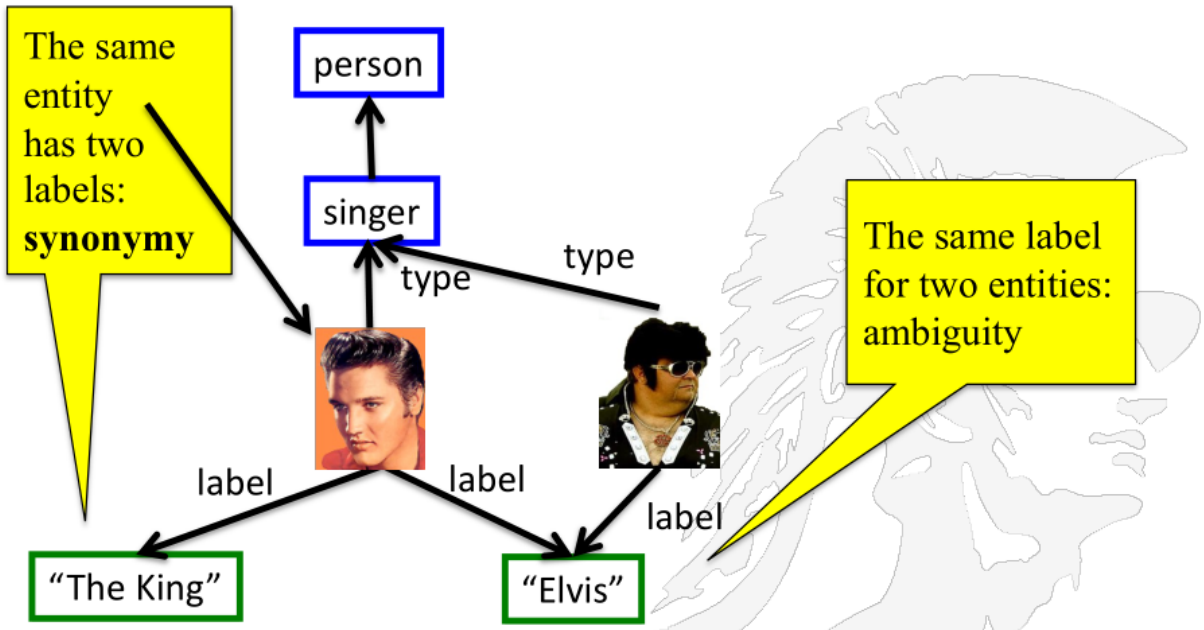
A knowledgebase can be seen as a directed labeled multigraph, where the nodes are entities and the edges relations.

A **multigraph** is a graph which is permitted to have multiple edges that have the same end nodes. Two vertices may be connected by more than one edge

University of Southern California

USC Viterbi School of Engineering

A Single Entity Can Have Different Labels

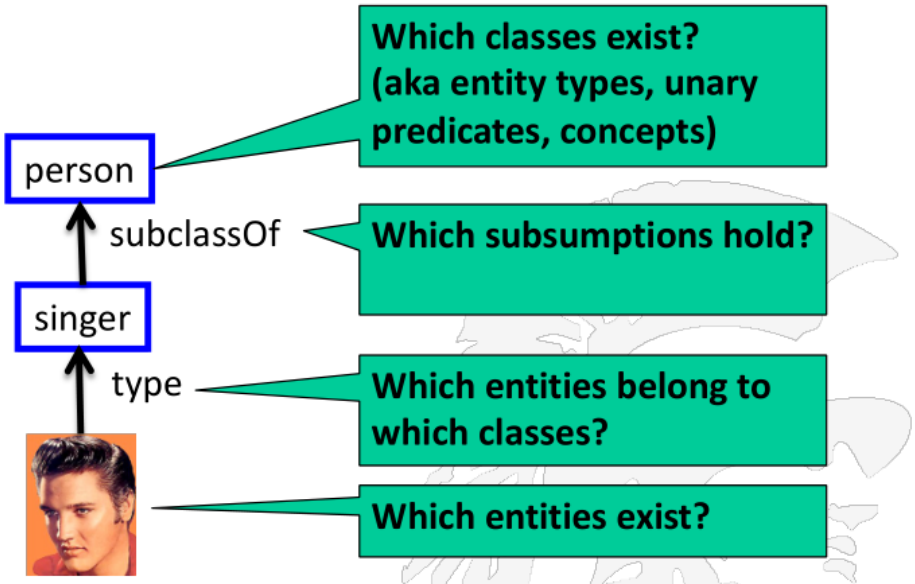


University of Southern California

USC Viterbi School of Engineering

USC

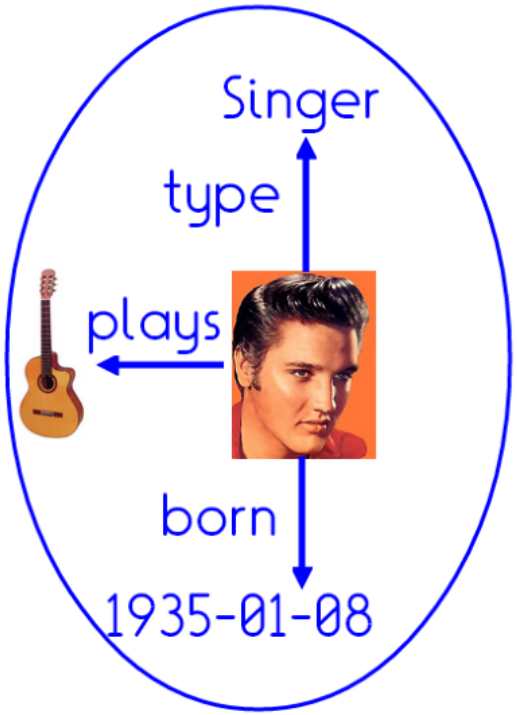
To Build a Knowledgebase One Must Find Classes and Instances



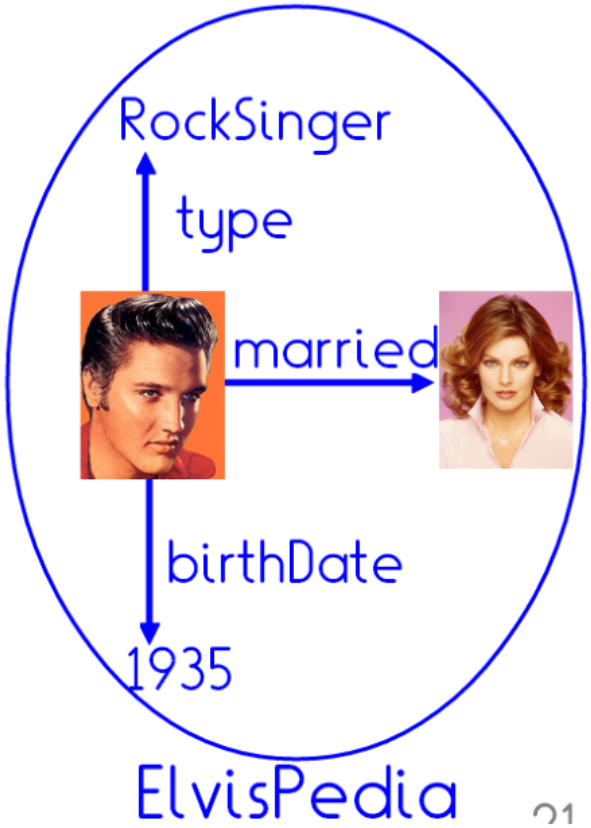
University of Southern California

USC Viterbi School of Engineering

Two Knowledgebases With Complementary Information



See https://en.wikipedia.org/wiki/YAGO_(database)



See https://theelvispedia.com/

21

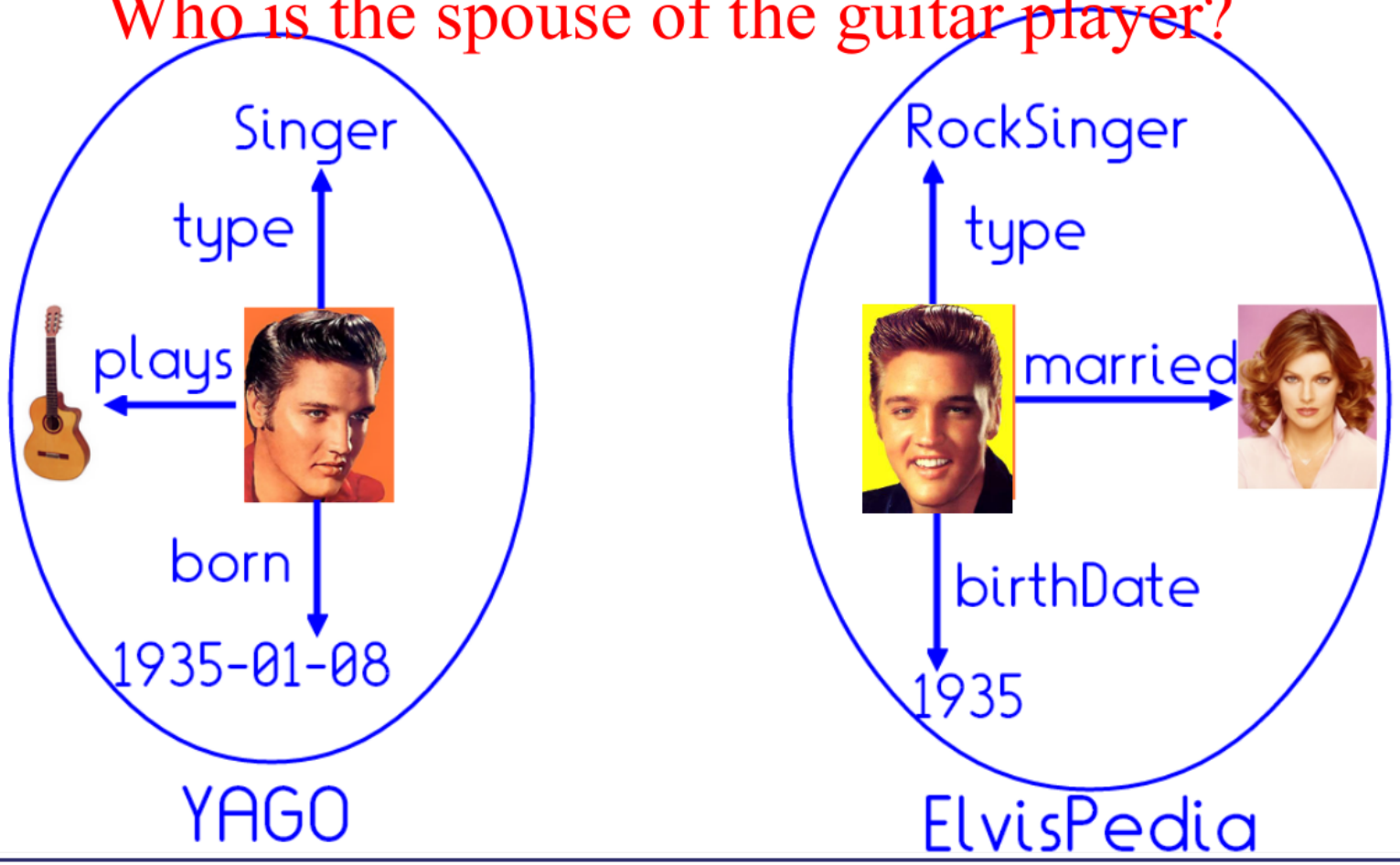
University of Southern California

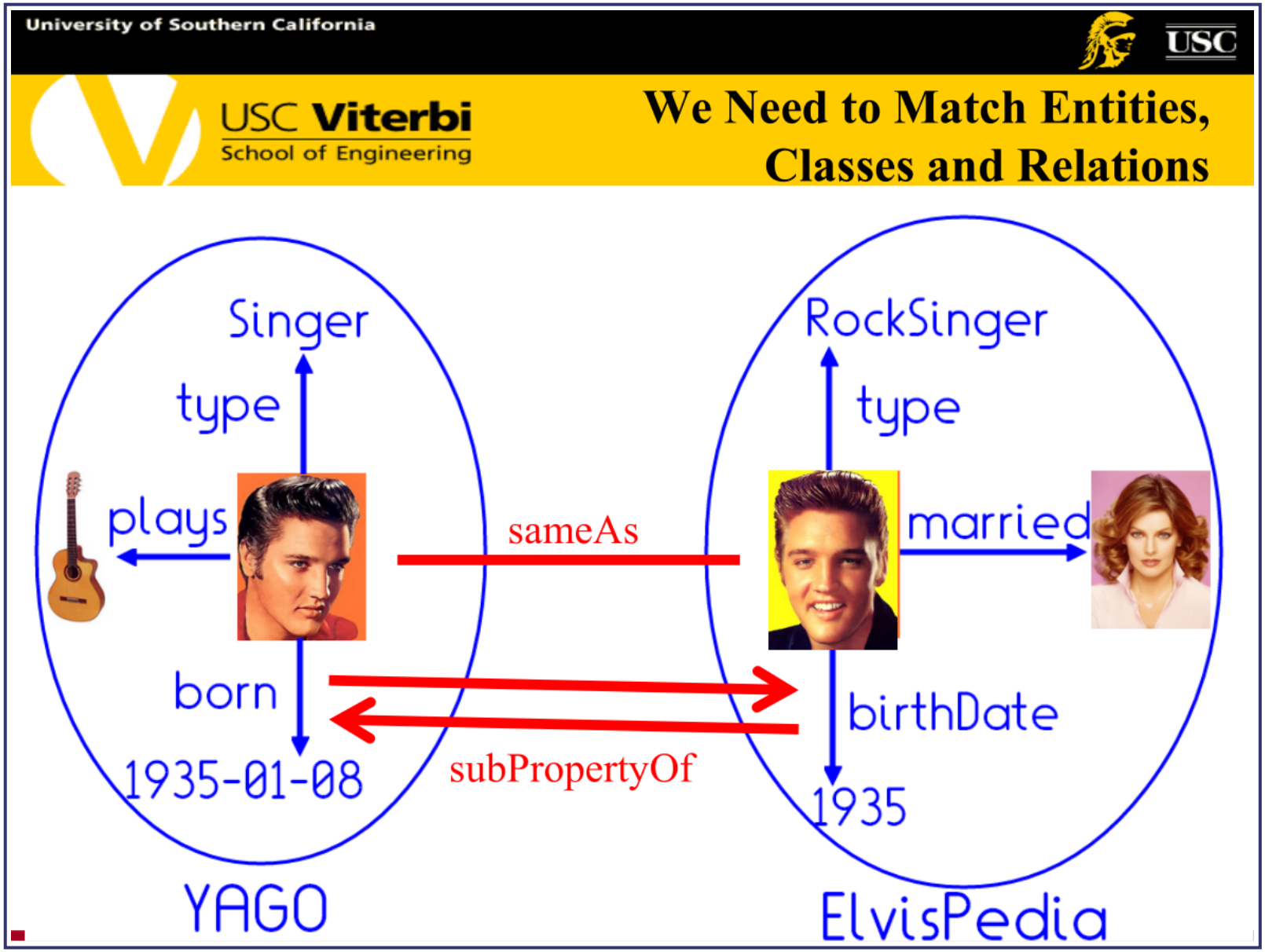
USC

USC Viterbi School of Engineering

A Knowledgebase Must Work Across Multiple Ontologies

Who is the spouse of the guitar player?




University of Southern California
USC
USC Viterbi
School of Engineering
We Need to Match Entities, Classes and Relations
Singer
type
plays
born
1935-01-08
YAGO
sameAs
subPropertyOf
RockSinger
type
married
birthDate
1935
ElvisPedia

University of Southern California

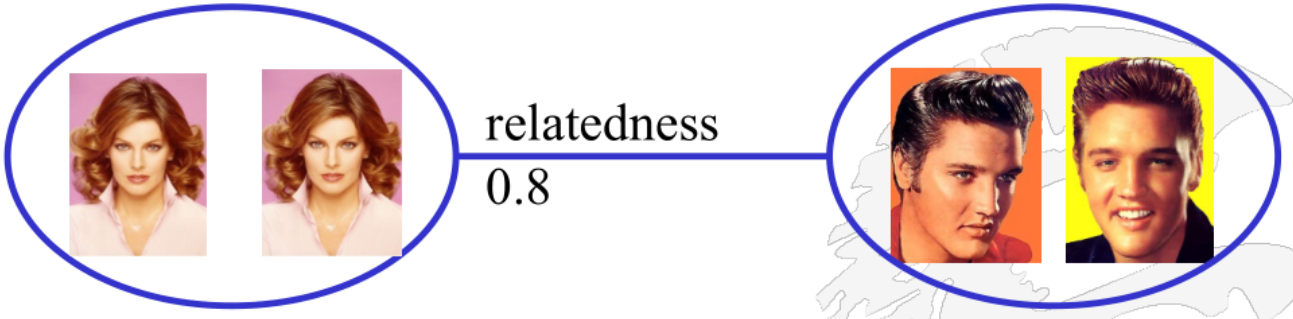
USC Viterbi School of Engineering

Combining Elements From Different Knowledgebases Means Matching Entities

Build a graph:

nodes: pairs of entities, weighted with similarity

edges: weighted with degree of relatedness



Iterate until convergence:

similarity := weighted sum of neighbor similarities

many variants (belief propagation, label propagation, etc.)

●●

University of Southern California

USC Viterbi School of Engineering

Inferencing on KnowledgeBases

- An **inference engine** is a component of a system that applies logical rules to a knowledgebase to deduce new information
- This process is ongoing as each new fact in the knowledgebase can trigger additional rules in the inference engine.
- Inference engines work primarily in one of two modes: forward chaining and backward chaining
  - **Forward chaining** starts with the known facts and asserts new facts.
  - **Backward chaining** starts with goals, and works backward to determine what facts must be asserted so that the goals can be achieved.
- An inference engine cycles through three sequential steps: match rules, select rules, and execute rules
- The execution of the rules will sometimes result in new facts or goals being added to the knowledgebase which will trigger the cycle to repeat. This cycle continues until no new rules can be matched
- Search engines typically use forward chaining

University of Southern California

USC Viterbi School of Engineering

Forward Chaining

- **Forward chaining** is the repeated application of modus ponens
- In propositional logic, ***modus ponens*** is the rule
  - "*P* implies *Q*" and "*P*" are both asserted to be true, so therefore *Q* must be true."
  - Sometimes modus ponens is written as ((P -> Q) and P) -> Q
  - For Example
    - **If today is Tuesday, then John will go to work.**
    - **Today is Tuesday.**
    - **Therefore, John will go to work.**
- Forward chaining starts with the available data and uses inference rules to extract more data until a goal or endpoint is reached.

University of Southern California

USC Viterbi School of Engineering

Binary Relations and Instances

Here are some sample binary relations with their type signature, e.g.

hasAdvisor: Person × Person
graduatedAt: Person × University
bornOn: Person × Date

Here are instances of the above binary relations

hasAdvisor (JimGray, MikeHarrison)
hasAdvisor (Susan Davidson, Hector Garcia-Molina)
graduatedAt (JimGray, Berkeley)
graduatedAt (HectorGarcia-Molina, Stanford)
bornOn (JohnLennon, 9-Oct-1940)

For more examples see
https://www.tutorialandexample.com/forward-chaining/
And
https://www.javatpoint.com/forward-chaining-and-backward-chaining-in-ai

27

••

University of Southern California

USC Viterbi School of Engineering

Semantic Network

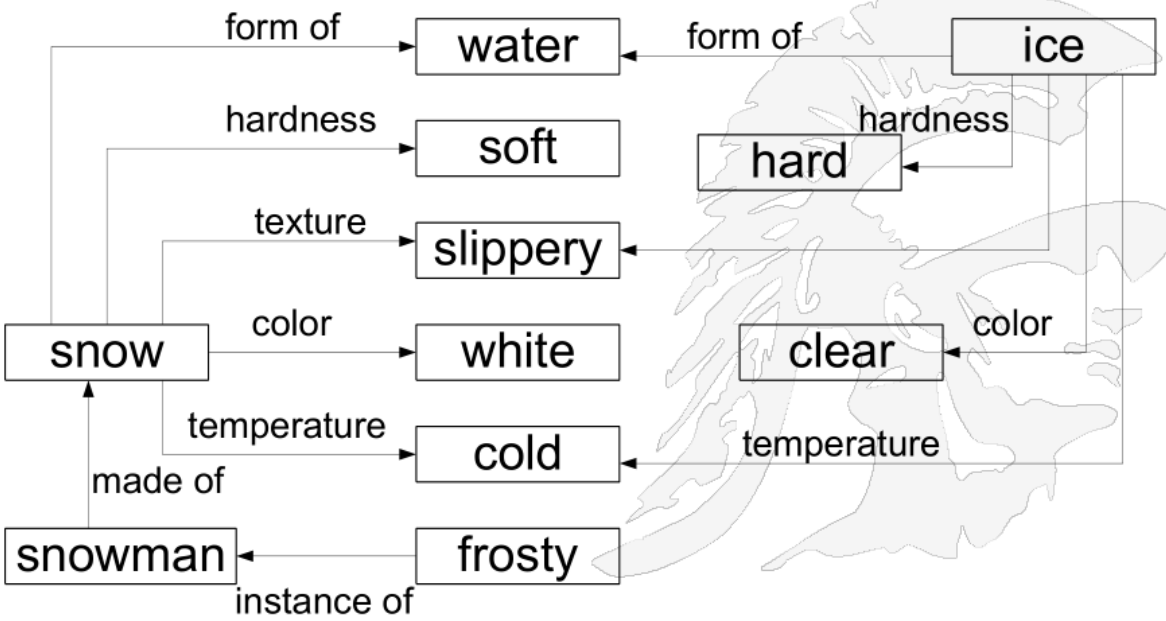
- A semantic network is a knowledge representation scheme that represents knowledge as a graph.
  - The nodes of the graph correspond to concepts or facts and
  - the arc correspond to relations or associations between concepts.
  - In a semantic network both the nodes and arcs are labeled

University of Southern California

USC Viterbi School of Engineering

Semantic Network

- A semantic network that defines the properties of snow and ice
- The concept snowman inherits all the properties of snow
- The concept of ice and snow share a number of properties



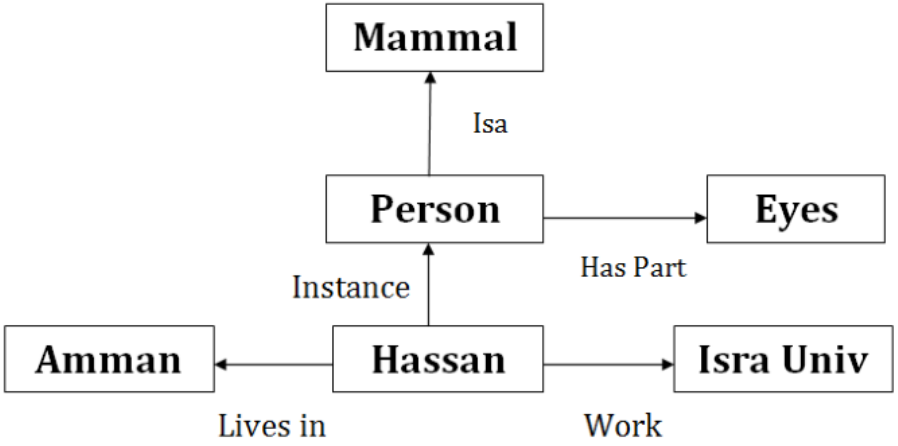
University of Southern California

USC Viterbi School of Engineering

USC

Expressing Relationship in Semantic Networks

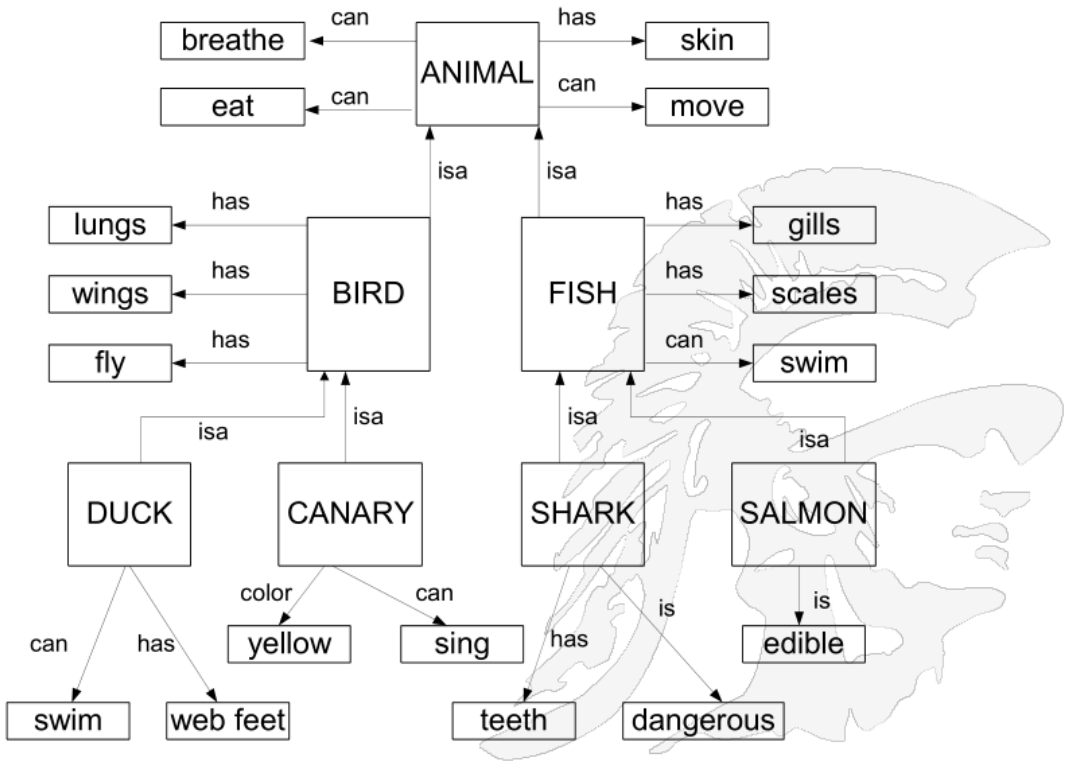
- The most important link in a semantic network is the is-a link
- A semantic network can organize knowledge in a hierarchy by using the is-a link such that the lower nodes inherit the properties of the higher nodes



University of Southern California

USC Viterbi School of Engineering

The is-a relationship in a Semantic Network



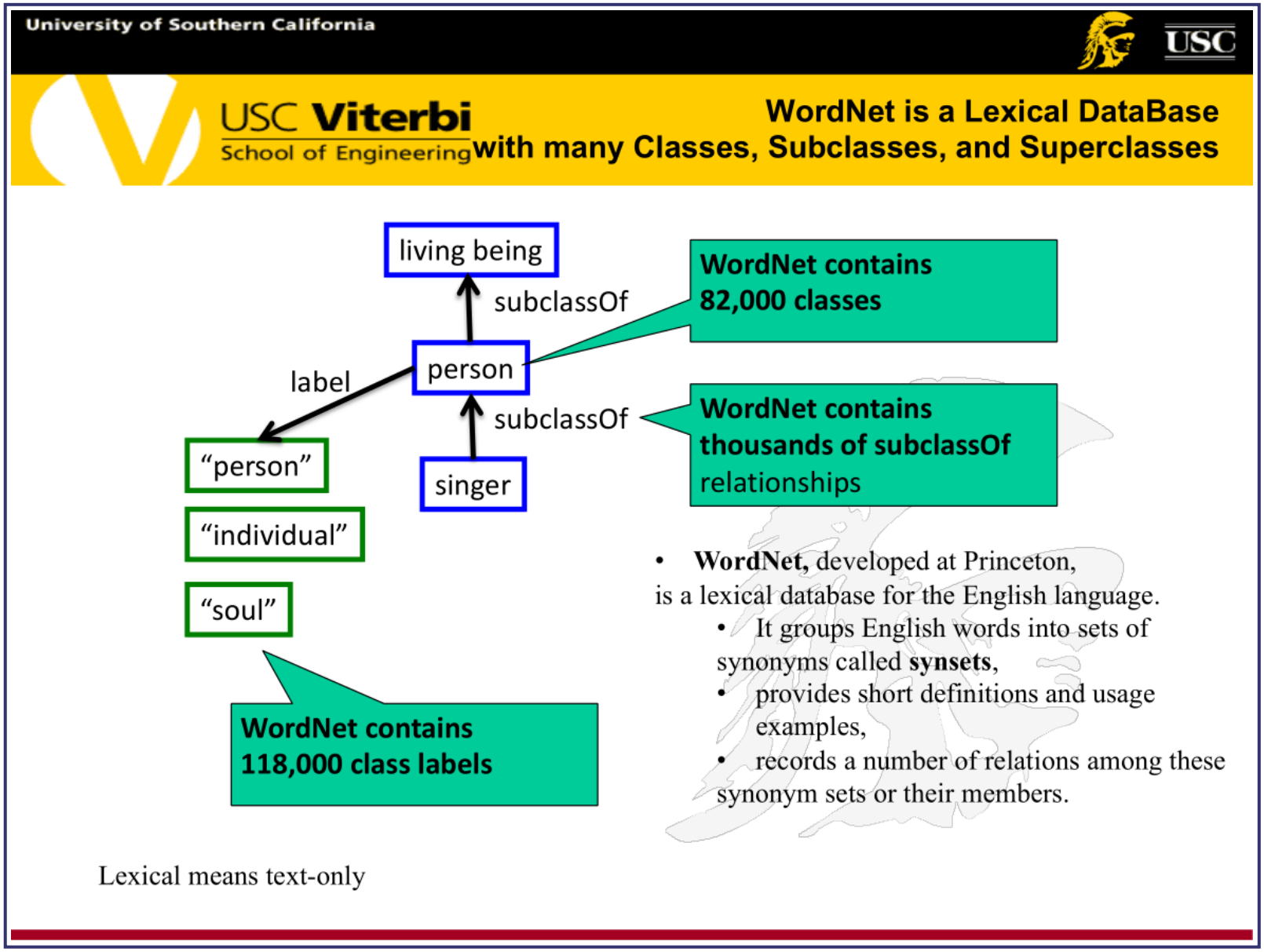
●●

University of Southern California

USC Viterbi
School of Engineering

WordNet

●●


University of Southern California
USC
USC Viterbi
School of Engineering
WordNet is a Lexical DataBase
with many Classes, Subclasses, and Superclasses
living being
subclassOf
person
label
subclassOf
"person"
singer
"individual"
"soul"
WordNet contains
82,000 classes
WordNet contains
thousands of subclassOf
relationships
WordNet contains
118,000 class labels
• WordNet, developed at Princeton,
is a lexical database for the English language.
• It groups English words into sets of synonyms called synsets,
• provides short definitions and usage examples,
• records a number of relations among these synonym sets or their members.
Lexical means text-only

●●

University of Southern California

USC Viterbi School of Engineering

WordNet Example: Superclass of Person

- S: (n) **person**, individual, someone, somebody, mortal, soul (a human being) *"there was too much for one person to do"*
  - *direct hyponym* / *full hyponym*
  - *part meronym*
  - *member holonym*
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
      - S: (n) living thing, animate thing (a living (or once living) entity)
        - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
          - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
            - S: (n) physical entity (an entity that has physical existence)
              - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Note the terms:

- **hyponym**
  - More specific
- **Holonym**
  - Denoting the whole
- **Hypernym**
  - A broad or superordinate

35

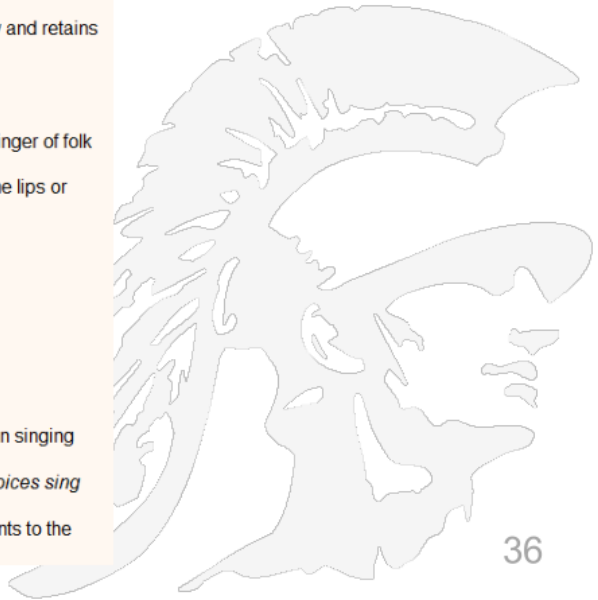
●●

University of Southern California

USC Viterbi School of Engineering

WordNet Example: Subclass of Singer

- S: (n) **singer**, vocalist, vocalizer, vocaliser (a person who sings)
  - ***direct hyponym*** / *full hyponym*
    - S: (n) alto (a singer whose voice lies in the alto clef)
    - S: (n) baritone, barytone (a male singer)
    - S: (n) bass, basso (an adult male singer with the lowest voice)
    - S: (n) canary (a female singer)
    - S: (n) caroler, caroller (a singer of carols)
    - S: (n) castrato (a male singer who was castrated before puberty and retains a soprano or alto voice)
    - S: (n) chorister (a singer in a choir)
    - S: (n) contralto (a woman singer having a contralto voice)
    - S: (n) crooner, balladeer (a singer of popular ballads)
    - S: (n) folk singer, jongleur, minstrel, poet-singer, troubadour (a singer of folk songs)
    - S: (n) hummer (a singer who produces a tune without opening the lips or forming words)
    - S: (n) lieder singer (a singer of lieder)
    - S: (n) madrigalist (a singer of madrigals)
    - S: (n) opera star, operatic star (singer of lead role in an opera)
    - S: (n) rapper (someone who performs rap music)
    - S: (n) rock star (a famous singer of rock music)
    - S: (n) songster (a person who sings)
    - S: (n) soprano (a female singer)
    - S: (n) tenor (an adult male with a tenor voice)
    - S: (n) thrush (a woman who sings popular songs)
    - S: (n) torch singer (a singer (usually a woman) who specializes in singing torch songs)
    - S: (n) voice ((metonymy) a singer) *"he wanted to hear trained voices sing it"*
    - S: (n) warbler (a singer; usually a singer who adds embellishments to the song)



36

●●

University of Southern California

USC Viterbi School of Engineering

WordNet Example: Instances But Very Few

- S: (n) **singer**, vocalist, vocalizer, vocaliser (a person who sings)
  - *direct hyponym* / *full hyponym*
  - ***has instance***
    - S: (n) Bailey, Pearl Bailey, Pearl Mae Bailey (United States singer (1918-1990))
    - S: (n) Cash, Johnny Cash, John Cash (United States country music singer and songwriter (1932-2003))
    - S: (n) Chevalier, Maurice Chevalier (French actor and cabaret singer (1888-1972))
    - S: (n) Dietrich, Marlene Dietrich, Maria Magdalene von Losch (United States film actress (born in Germany) who made many films with Josef von Sternberg and later was a successful cabaret star (1901-1992))
    - S: (n) Dylan, Bob Dylan (United States songwriter noted for his protest songs (born in 1941))
    - S: (n) Fitzgerald, Ella Fitzgerald (United States scat singer (1917-1996))
    - S: (n) Garland, Judy Garland (United States singer and film actress (1922-1969))
    - S: (n) Horne, Lena Horne, Lena Calhoun Horne (United States singer and actress (born in 1917))
    - S: (n) Iglesias, Julio Iglesias (Spanish singer noted for his ballads and love songs (born in 1943))
    - S: (n) Jackson, Mahalia Jackson (United States singer who did much to popularize gospel music (1911-1972))
    - S: (n) Jackson, Michael Jackson, Michael Joe Jackson (United States singer who began singing with his four brothers and later became a highly successful star during the 1980s (born in 1958))

only 32 singers !?
4 guitarists
5 scientists
0 enterprises
2 entrepreneurs

WordNet classes lack instances

37

●●

University of Southern California

USC

USC Viterbi
School of Engineering

Wikipedia

••

University of Southern California

USC Viterbi School of Engineering

Wikipedia: Tranformation from Database to KnowledgeBase

Wikipedia's Original Mission Statement
"Imagine a world in which every person on the planet shares in the sum of all human knowledge. That is what we're doing.
(for free, in the language of their choice)

••

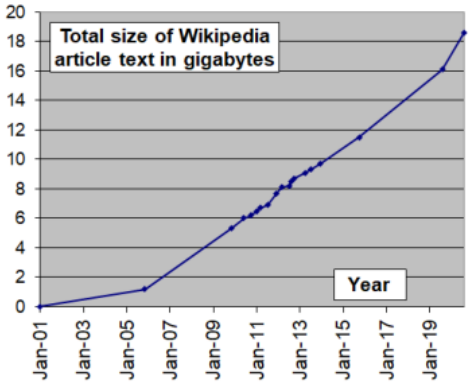
University of Southern California

USC Viterbi School of Engineering

Wikipedia's Scale

- As of April, 2019 Wikipedia's database when dumped takes up 100GBs compressed, 10TBs uncompressed
- Total wiki pages: 51,000,000+
- Total English articles: 6.1 million
- Unique visitors per month: 500 million
- Monthly mobile page views: 3.7 billion

Jimmy Wales, Founder



By Mikael Häggström ,

●●

University of Southern California

USC Viterbi School of Engineering

Wikipedia's Five Pillars (5P)

1. **Encyclopaedia**
   - Notable topics
   - No original research (NOR)
2. **Neutral point of view (NPOV)**
   - Verifiability (referencing)
3. **Free content**
   - Anyone can edit
   - No copyright infringements
4. **Be civil**
5. **No firm rules**

Wikipedia:Five pillars

From Wikipedia, the free encyclopedia

The fundamental principles by which Wikipedia operates can be summarized in the form of five "pillars":

Wikipedia is an encyclopedia.

It incorporates elements of general and specialized encyclopedias, almanacs, and gazetteers. Wikipedia is not a soapbox, an advertising platform, a vanity press, an experiment in anarchy or democracy, an indiscriminate collection of information, or a web directory. It is not a dictionary, a newspaper, or a collection of source documents; that kind of content should be contributed instead to the Wikimedia sister projects.

Wikipedia is written from a neutral point of view.

We strive for articles that document and explain the major points of view in a balanced and impartial manner. We avoid advocacy and we characterize information and issues rather than debate them. In some areas there may be just one well-recognized point of view; in other areas we describe multiple points of view, presenting each accurately and in context, and not presenting any point of view as "the truth" or "the best view". All articles must strive for verifiable accuracy; unreferenced material may be removed, so please provide references. Editors' personal experiences, interpretations, or opinions do not belong here. That means citing verifiable, authoritative sources, especially on controversial topics and when the subject is a living person.

Wikipedia is free content that anyone can edit, use, modify, and distribute.

Respect copyright laws, and do not plagiarize sources. Non-free content is allowed under fair use, but strive to find free alternatives to any media or content that you wish to add to Wikipedia. Since all your contributions are freely licensed to the public, no editor owns any article; all of your contributions can and will be mercilessly edited and redistributed.

Editors should interact with each other in a respectful and civil manner.

Respect and be polite to your fellow Wikipedians, even when you disagree. Apply Wikipedia etiquette, and avoid personal attacks. Find consensus, avoid edit wars, and remember that there are 4,143,499 articles on the English Wikipedia to work on and discuss. Act in good faith, and never disrupt Wikipedia to illustrate a point. Be open and welcoming, and assume good faith on the part of others. When conflict arises, discuss details on the talk page, and follow dispute resolution.

Wikipedia does not have firm rules.

Rules in Wikipedia are not carved in stone, as their wording and interpretation are likely to change over time. The principles and spirit of Wikipedia's rules matter more than their literal wording, and sometimes improving Wikipedia requires making an exception to a rule. Be bold (but not reckless) in updating articles and do not worry about making mistakes. Prior versions of pages are saved, so any mistakes can be corrected.

Wikipedia principles [show]

Key Wikipedia policies and guidelines [show]

Category (+/−): Wikipedia basic information (−) (±) | (+)

This page was last modified on 15 December 2012, at 03:41.

Text is available under the Creative Commons Attribution/Share-Alike License; additional terms may apply. See Terms of use for details.

••

University of Southern California

USC Viterbi School of Engineering

Wikipedia Statistics

- Among top 10 most visited websites
- 70% of traffic is from search engines
- Cited in hundreds of U.S. court rulings

https://stats.wikimedia.org/EN/Sitemap.htm
and
https://stats.wikimedia.org/#/all-projects

●●

University of Southern California

USC Viterbi School of Engineering

Wikipedia is a Rich Source of Instances

Wikipedia founders →

Jimmy Wales

Larry Sanger

Steve Jobs

From Wikipedia, the free encyclopedia

For the biography, see Steve Jobs (biography).

Steven Paul Jobs (/ˈdʒɒbz/; February 24, 1955 – October 5, 2011)[4][5] was an American businessman and inventor widely recognized as a charismatic pioneer of the personal computer revolution.[6][7] He was co-founder, chairman, and chief executive officer of Apple Inc. Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, following the acquisition of Pixar by Disney.

In the late 1970s, Apple co-founder Steve Wozniak engineered one of the first commercially successful lines of personal computers, the Apple II series. Jobs directed its aesthetic design and marketing along with A.C. "Mike" Markkula, Jr. and others. In the early 1980s, Jobs was among the first to see the commercial potential of Xerox PARC's mouse-driven graphical user interface, which led to the creation of the Apple Lisa (engineered by Ken Rothmuller and John Couch) and, one year later, creation of Apple employee Jef Raskin's Macintosh.

After losing a power struggle with the board of directors in 1985, Jobs left Apple and founded NeXT, a computer platform development company specializing in the higher-education and business markets. NeXT was eventually acquired by Apple in 1996, which brought Jobs back to the company he co-founded, and provided Apple with the NeXTSTEP codebase, from which the Mac OS X was developed."[8] Jobs was named Apple advisor in 1996, interim CEO in 1997, and CEO from 2000 until his resignation. He oversaw the development of the iMac, iTunes, iPod, iPhone, and iPad and the company's Apple Retail Stores.[9] In 1986, he acquired the computer graphics division of Lucasfilm Ltd, which was spun off as Pixar Animation Studios.[10] He was credited in *Toy Story* (1995) as an executive producer. He remained CEO and majority shareholder at 50.1 percent until its acquisition by The Walt Disney Company in 2006,[11] making Jobs Disney's largest individual shareholder at seven percent and a member of Disney's Board of Directors.[12][13]

In 2003, Jobs was diagnosed with a pancreas neuroendocrine tumor. Though it was initially treated, he reported a hormone imbalance, underwent a liver transplant in 2009, and appeared progressively thinner as his health declined.[14] On medical leave for most of 2011, Jobs resigned as Apple CEO in August that year and was elected Chairman of the Board. On October 5, 2011, Jobs died of respiratory arrest related to his metastatic tumor. He

Steve Jobs

Jobs holding a white iPhone 4 at Worldwide Developers Conference 2010

Born	Steven Paul Jobs February 24, 1955[1][2] San Francisco, California, U.S.[1][2]
Died	October 5, 2011 (aged 56)[2] Palo Alto, California, U.S.
Nationality	American
Alma mater	Reed College (dropped out)

43

●●

University of Southern California

USC Viterbi School of Engineering

Wikipedia's Categories Also Contain Classes

Categories: Steve Jobs | 1955 births | 2011 deaths | American adoptees | American billionaires | American chief executives | American computer businesspeople | American industrial designers | American inventors | American people of German descent | American people of Swiss descent | American people of Syrian descent | American technology company founders | American Zen Buddhists | Apple Inc. | Apple Inc. employees | Businesspeople from California | Businesspeople in software | Cancer deaths in California | Computer designers | Computer pioneers | Deaths from pancreatic cancer | Disney people | Internet pioneers | National Medal of Technology recipients | NeXT | Organ transplant recipients | People from the San Francisco Bay Area | Pescetarians | Reed College alumni

But categories do not form a taxonomic hierarchy, i.e. there is no ISA hierarchy

An isa hierarchy only specifies that a set of objects is the subclasses of another object, but nothing more

44

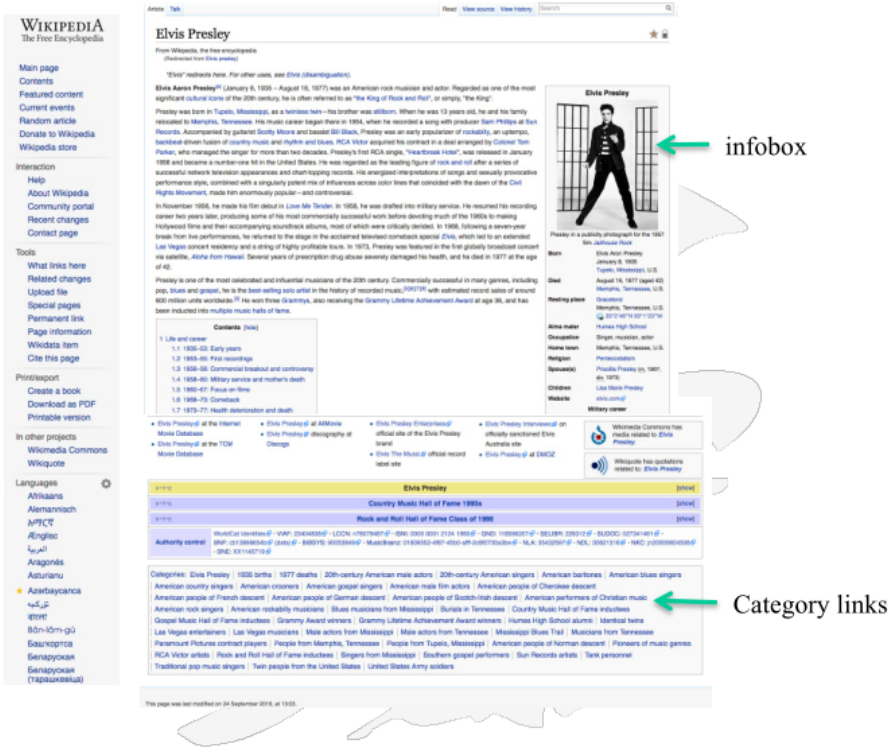
●●

University of Southern California

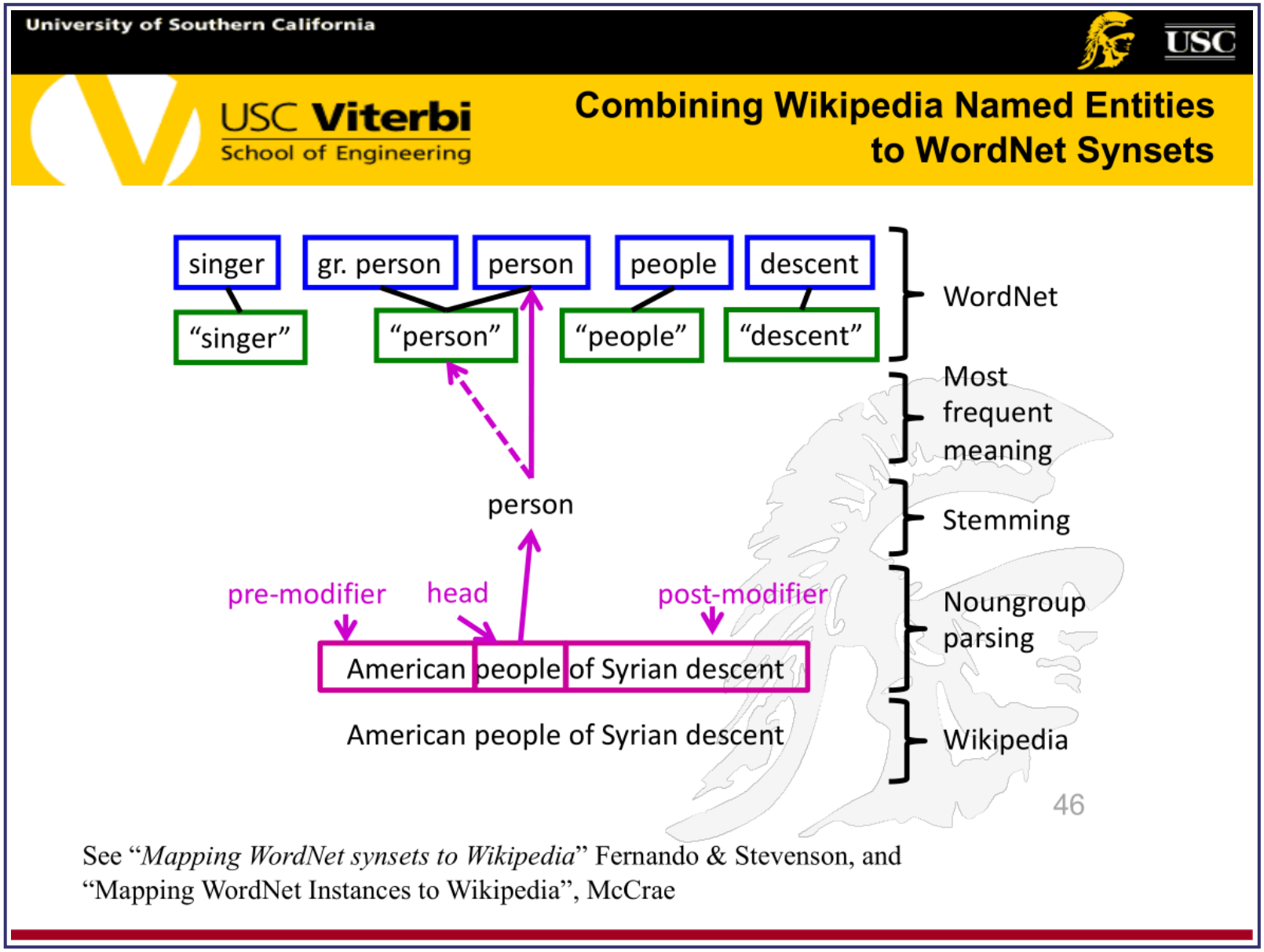
USC Viterbi School of Engineering

Structure of a Wikipedia Page

- **Types of links**
  - ***Article links***
    - **links from one article to another of the same language;**
  - ***Category links***
    - **links from an article to special "Category" pages;**
  - ***Interlingual links***
    - **links from an article to a presumably equivalent, article in another language;**
- **Types of special pages**
  - ***Redirect pages***
    - **short pages which often provide equivalent names for an entity**
  - ***Disambiguation pages***
    - **a page with little content that links to multiple similarly named articles.**
- **Infoboxes, templates, list pages, wikipedia commons, ...**



●●


University of Southern California
USC
USC Viterbi
School of Engineering
Combining Wikipedia Named Entities to WordNet Synsets
singer
gr. person
person
people
descent
"singer"
"person"
"people"
"descent"
WordNet
Most frequent meaning
person
Stemming
pre-modifier
head
post-modifier
American people of Syrian descent
Noungroup parsing
American people of Syrian descent
Wikipedia
46
See "Mapping WordNet synsets to Wikipedia" Fernando & Stevenson, and "Mapping WordNet Instances to Wikipedia", McCrae

••

University of Southern California

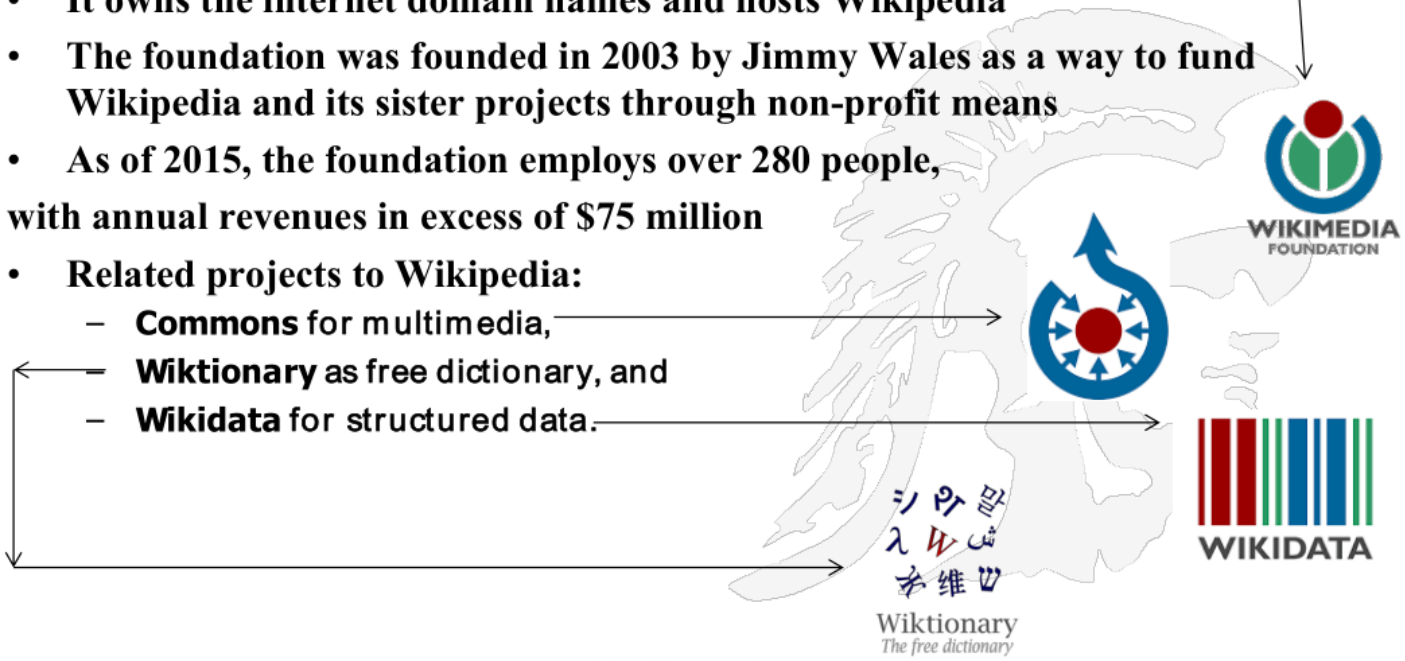
USC Viterbi School of Engineering

Wikimedia Foundation

- Wikimedia Foundation, Inc. (WMF) is an American non-profit and charitable organization headquartered in San Francisco.
- It owns the internet domain names and hosts Wikipedia
- The foundation was founded in 2003 by Jimmy Wales as a way to fund Wikipedia and its sister projects through non-profit means
- As of 2015, the foundation employs over 280 people,

with annual revenues in excess of $75 million

- Related projects to Wikipedia:
  - **Commons** for multimedia,
  - **Wiktionary** as free dictionary, and
  - **Wikidata** for structured data.



 47

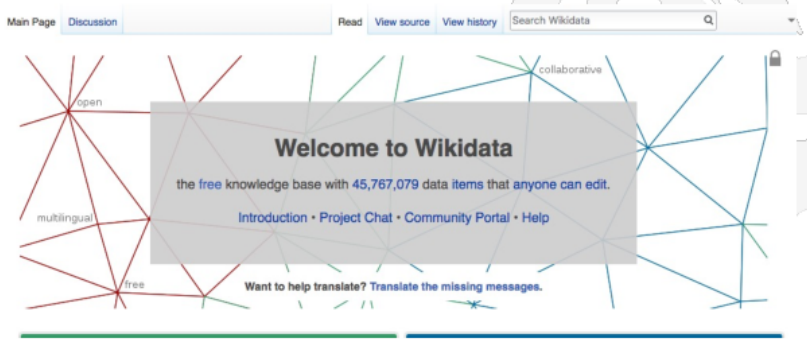
••

University of Southern California

USC Viterbi School of Engineering

WikiData

- WikiData is an effort to convert the Wikipedia data into a knowledgebase
- WikiData aims to create a free RDF-like KB about the world that can be read/edited by humans & machines
- Wikidata clients use the repository, e.g. to populate Web pages or Wikipedia infoboxes
- WikiData increases the quality and lowers the maintenance costs of Wikipedia and related projects



48

●●

University of Southern California

USC Viterbi School of Engineering

Wikidata Multilingual Coverage

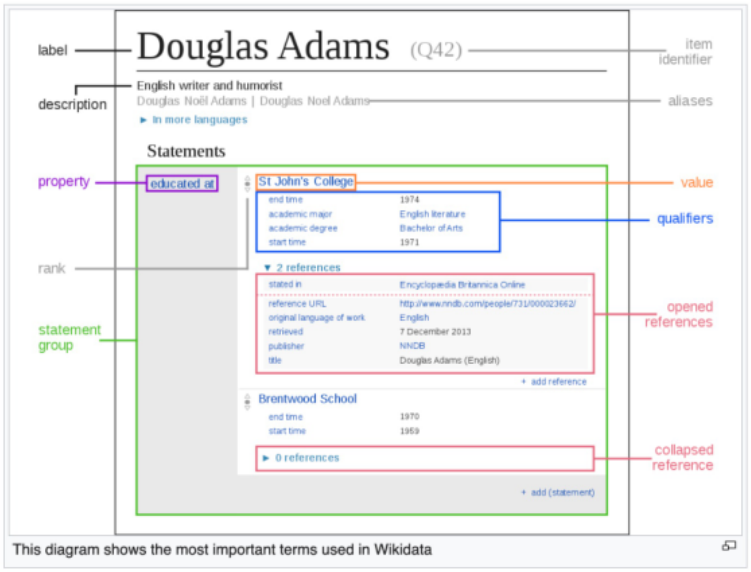
- → The challenge: Wikipedia has many named entities that occur in numerous articles
  - → E.g Ex-President Obama or President Trump are mentioned in over 100 articles
  - → If one of them dies, this must be reflected on many of those pages
- → Solution: Let the entry for Obama/Trump be centered in Wikidata and let all references to Obama/Trump point to the Wikidata entry
- → Another aspect of Wikidata is their multilingual coverage
  - → Popular entities are present in many languages (up to 180); and even in one Wikipedia page there may be many languages
  - → E.g. Lucas Cranach (German Renaissance Painter) in Wikidata is referenced in 57 language tags, representing 44 languages and 13 language variants

●●

University of Southern California

USC Viterbi School of Engineering

Wikipedia Page and WikiData Page for Douglas Adams



His Wikipedia page His WikiData page

Statements in Wikidata consist of key-value pairs, which match a **property** (such as "author", or "publication date") with one or more **values** (such as “Sir Arthur Conan Doyle” or "1902").

●●

University of Southern California

USC

USC Viterbi
School of Engineering

Google's Knowledge Graph

●●

University of Southern California

USC Viterbi School of Engineering

Google Knowledge Graph

- Introduced in 2012 with the Hummingbird update, see https://searchengineland.com/google-hummingbird-172816
- Powered in part by Freebase
- KnowledgeGraph was accused of taking away traffic from Wikipedia
- Knowledge panels are information boxes for entities (person, place, organization, event, etc)
- A common source of info is Wikipedia, LinkedIn, Crunchbase, Reuters, Bloomberg

https://www.google.com/intl/es419/insidesearch/features/search/knowledge.html

Google's slogan for the knowledge graph: "things, not strings"

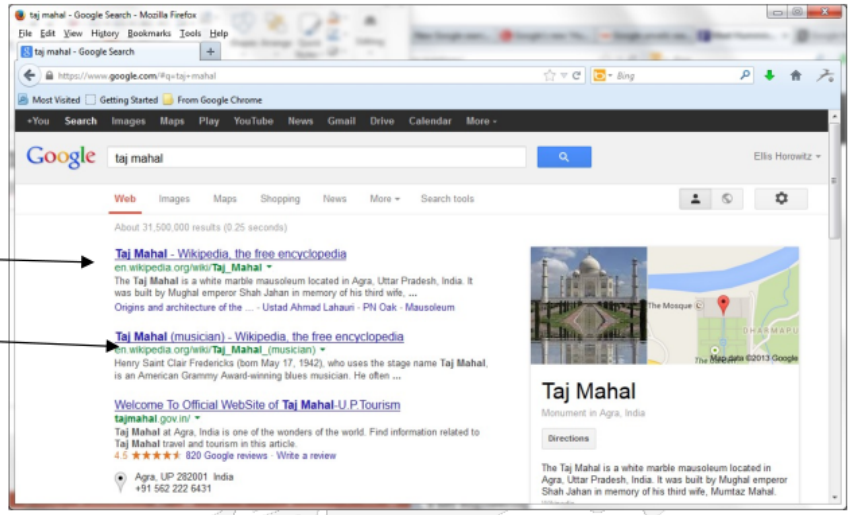
https://www.youtube.com/watch?v=mmQl6VGvX-c

●●

University of Southern California

USC Viterbi School of Engineering

Knowledge Graph Enhances Google Search in 3 main ways (1):



1. To improve the variety of search results, Google uses the knowledge graph to locate alternate interpretations of query terms,

Here it offers two of them with the same name e.g.
"taj mahal" - the mausoleum or musician

 53

●●

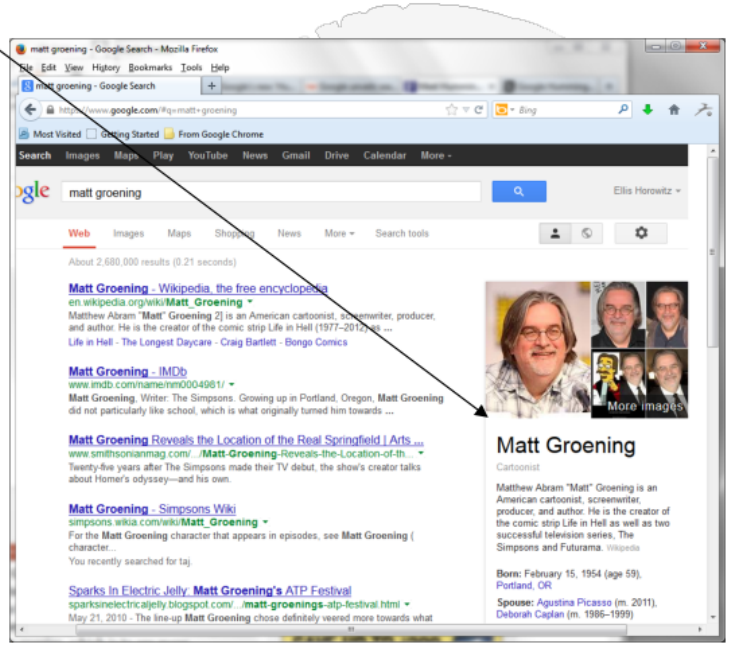
University of Southern California

USC Viterbi School of Engineering

Knowledge Graph Enhances Google Search in 3 main ways (2):

2. To provide deeper and broader results, typically in an info box
e.g. person entities include relations such as age, birthplace, marital status, children, education, etc.,
here is a sample result for Matt Groening

- creator of The Simpsons
- **Go Deeper**
  - his photo
  - when he was born
  - his spouse
  - his parents
  - why he is famous
- **Go Broader**
  - other people related to Groening



●●

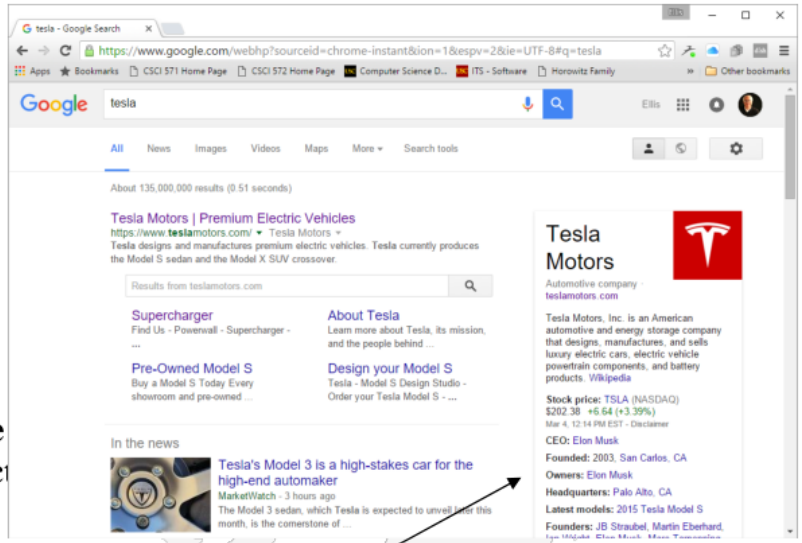
University of Southern California

USC Viterbi School of Engineering

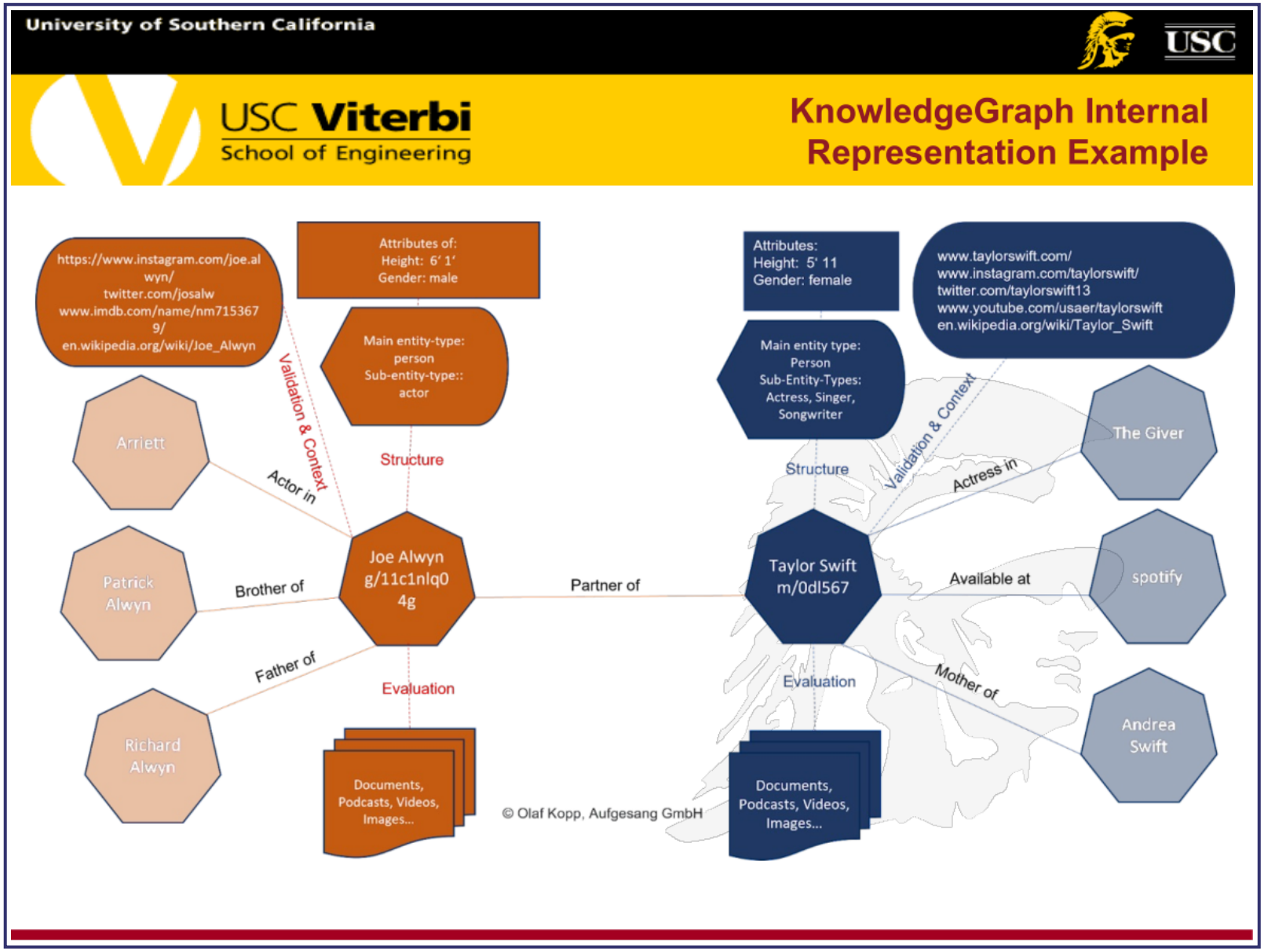
Knowledge Graph Enhances Google Search in 3 main ways: (3)

3. To provide the best summary
the knowledge graph exploits the relationships among the entities
e.g. the query "Tesla"

The knowledge graph allows Google to summarize relevant content around that topic, including key facts you're likely to need for that particular thing. E.g.
Tesla Motors, Inc. is an American automotive and energy storage company that designs, manufactures, and sells luxury electric cars, electric vehicle powertrain components, and battery produc



●●

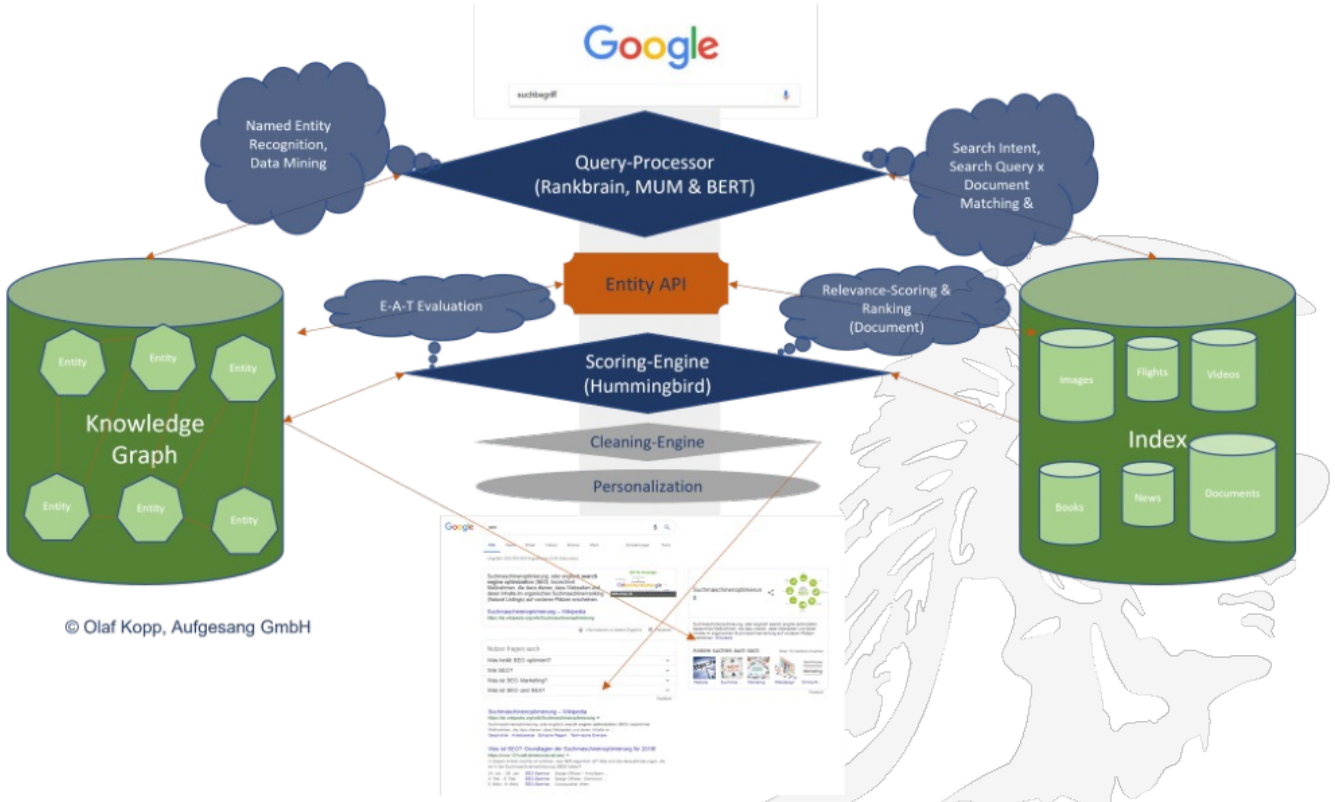

University of Southern California
USC
USC Viterbi
School of Engineering
KnowledgeGraph Internal Representation Example
https://www.instagram.com/joe.alwyn/
twitter.com/josalw
www.imdb.com/name/nm7153679/
en.wikipedia.org/wiki/Joe_Alwyn
Attributes of:
Height: 6' 1'
Gender: male
Main entity-type:
person
Sub-entity-type::
actor
Validation & Context
Structure
Arriett
Actor in
Patrick Alwyn
Brother of
Joe Alwyn
g/11c1nlq04g
Father of
Evaluation
Richard Alwyn
Documents, Podcasts, Videos, Images...
Partner of
Attributes:
Height: 5' 11
Gender: female
Main entity type:
Person
Sub-Entity-Types:
Actress, Singer, Songwriter
www.taylorswift.com/
www.instagram.com/taylorswift/
twitter.com/taylorswift13
www.youtube.com/usaer/taylorswift
en.wikipedia.org/wiki/Taylor_Swift
Validation & Context
Structure
Actress in
The Giver
Taylor Swift
m/0dl567
Available at
spotify
Evaluation
Mother of
Andrea Swift
Documents, Podcasts, Videos, Images...
© Olaf Kopp, Aufgesang GmbH

●●

University of Southern California

USC Viterbi School of Engineering

Google Search Combines Document Index with Knowledge Graph



When Google recognizes in the search query that it is about an entity recorded in the KnowledgeGraph it focuses its search around the entity