Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 23

Nov 3, 2017

Slide credit: Probase Microsoft Research Asia, YAGO Max Planck Institute, National Lib. Of Medicine, NIH

NLP Practical Goal for FOL: the ultimate Web question-answering system?

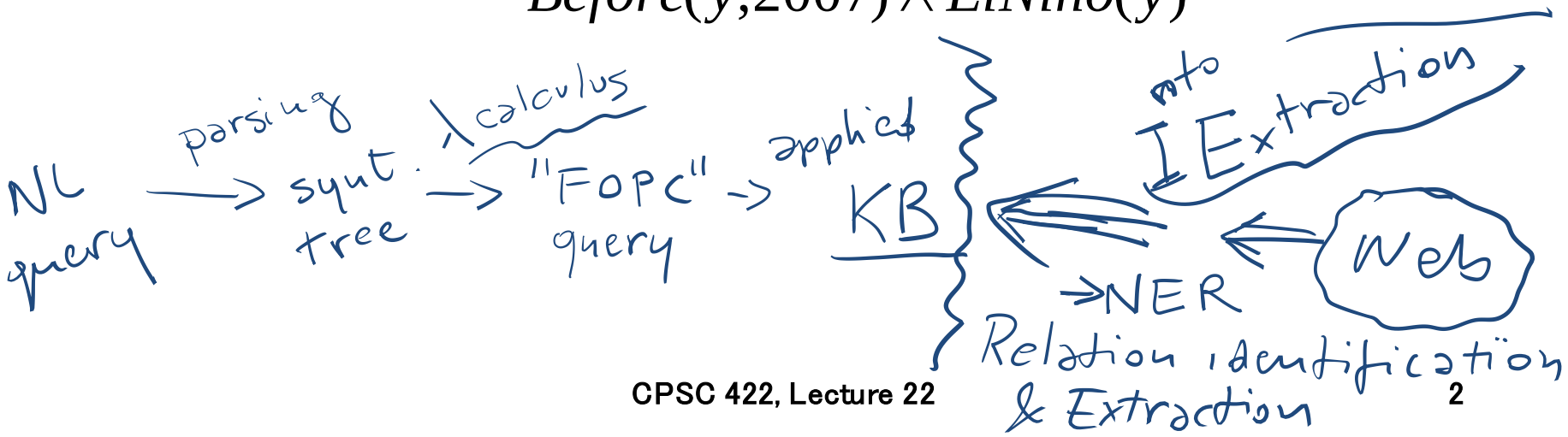
Map NL queries into FOPC so that answers can be effectively computed

What African countries are not on the Mediterranean Sea?

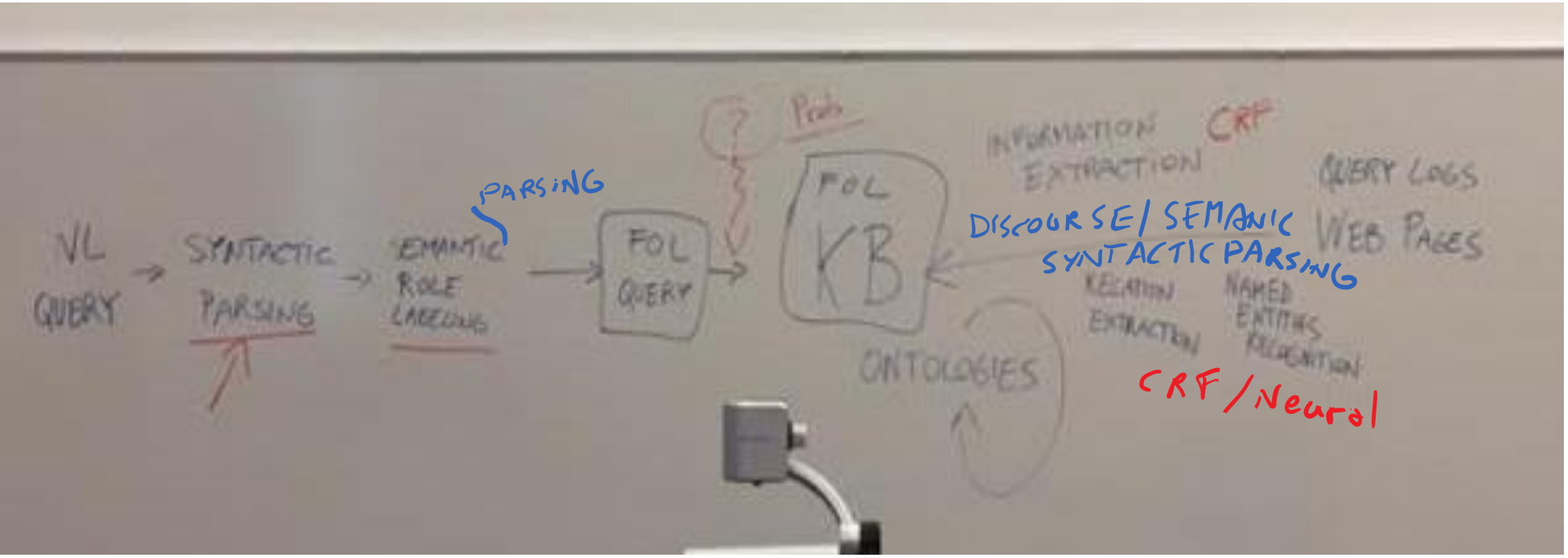
$$\exists c\ Country(c) \wedge \neg Borders(c, Med.Sea) \wedge In(c, Africa)$$

- *Was 2007 the first El Nino year after 2001?*

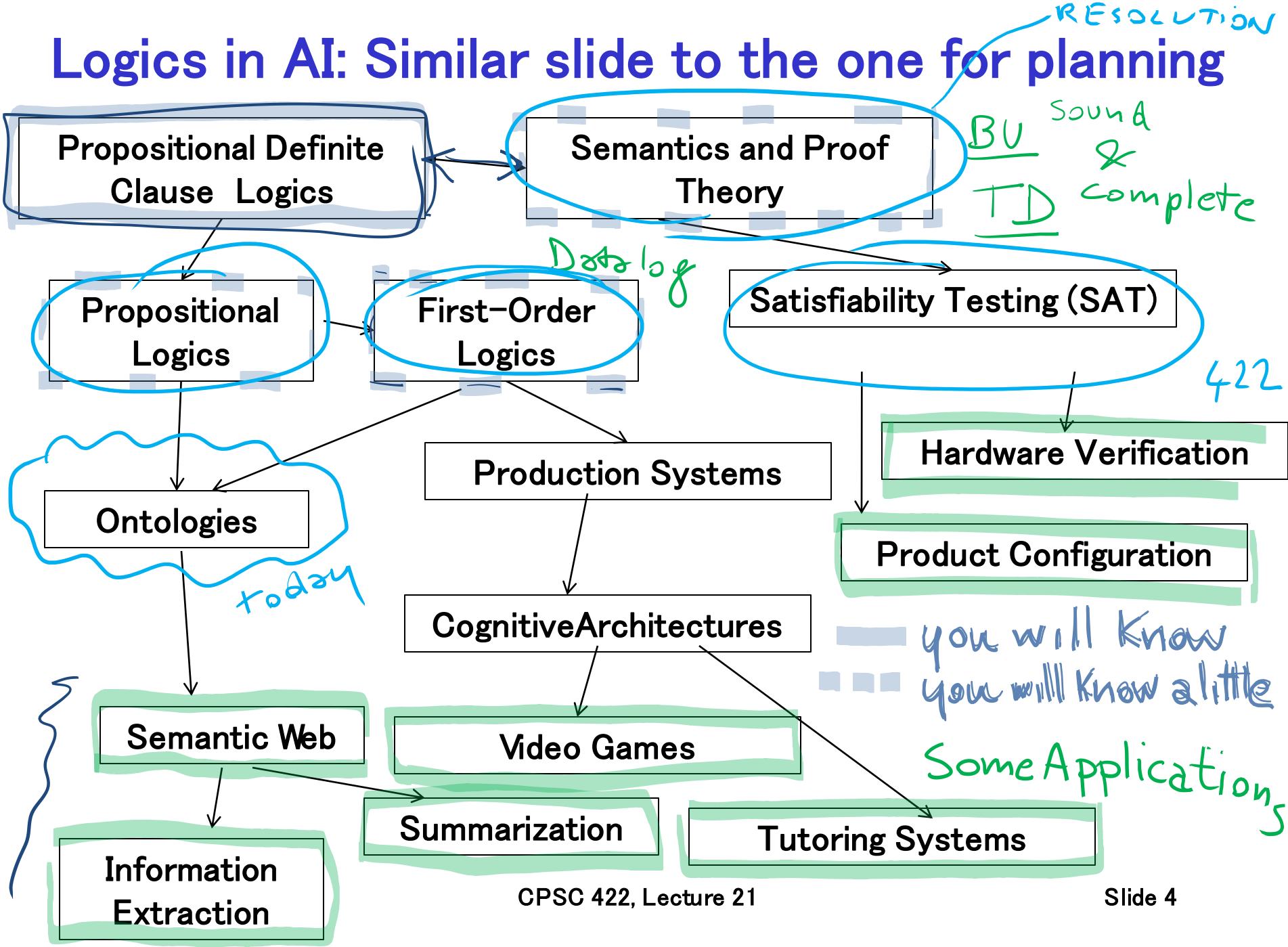
$$ElNino(2007) \wedge \neg \exists y\ Year(y) \wedge After(y, 2001) \wedge Before(y, 2007) \wedge ElNino(y)$$



Just a sketch: to provide some context for some concepts / techniques cobvered in 422



Logics in AI: Similar slide to the one for planning



Lecture Overview

EACH CAN BE DOWNLOADED

- **Ontologies** – what objects/individuals should we represent? what relations (unary, binary,..)?
- Inspiration from **Natural Language**: WordNet and FrameNet
- Extensions based on Wikipedia and mining the Web (YAGO, ProBase, Freebase)
- Domain Specific Ontologies (e.g., Medicine: MeSH, UMLS)

CHECK WEB INTERFACES FOR EACH OF THE ABOVE LINKS ON THE COURSE WEB PAGE

Ontologies

Given a logical representation (e.g., FOL)

What individuals and relations are there and we need to model?

In **AI** an **Ontology** is a specification of what **individuals** and **relationships** are **assumed to exist** and what terminology is used for them

- What **types** of individuals
- What **properties** of the individuals

Ontologies: inspiration from Natural Language

:

How do we refer to individuals and relationship in the world in Natural Languages e.g., English? Words

Where do we find definitions for words? Dictionary

Most of the definitions are circular? They are descriptions.

Red / Blood

Fortunately, there is still some useful semantic info (Lexical Relations):

- w_1 w_2 same Form and Sound, different Meaning — Homonymy (plant / bat)
- w_1 w_2 same Meaning, different Form — Synonymy (big/large)
- w_1 w_2 "opposite" Meaning — Antonymy (good/bad)
- w_1 w_2 $Meaning_1$ subclass of $Meaning_2$ — Hyponymy (dog/animal)

Polysemy

Def. The case where we have a set of words with the same form and multiple related meanings.

Consider the homonym:

bank → commercial **bank**$_1$ vs. river **bank**$_2$

- **Now consider:** "VGH is the hospital with the largest blood *bank* in BC" or
- "*A PCFG can be trained using derivation trees from a tree bank annotated by human experts*"

- **Are these a new independent senses of bank?**

Synonyms

Def. Different words with the same meaning.

Substitutability- if they can be substituted for one another in *some* environment without changing meaning or acceptability.

Would I be flying on a large/big plane?

?... became kind of a large/big sister to...

? You made a large/big mistake

Hyponymy/Hypernym

Def. Pairings where one word denotes a sub/super class of the other

- Since dogs are canids
  - ✓Dog is a *hyponym* of canid and
  - ✓Canid is a *hypernym* of dog

car/vehicle

doctor/human

……

Lexical Resources

Databases containing all lexical relations among all words

- **Development:**
  - **Mining info from dictionaries and thesauri**
  - **Handcrafting it from scratch**
- **WordNet: first developed with reasonable coverage and widely used, started with [Fellbaum… 1998]**
  - **for English (versions for other languages have been developed – see MultiWordNet)**

WordNet 3.0

Part Of Speech	Unique Strings	Word-Sense Pairs	Synsets
Noun	117798	146312	82115
Verb	11529	25047	13767
Adjective	21479	30002	18156
Adverb	4481	5580	3621
Totals	155287	206941	117659

- **For each word: all possible senses** (no distinction between homonymy and polysemy)
- **For each sense:** a set of synonyms (synset) and a *gloss*

WordNet: entry for "table"

The noun "table" has 6 senses in WordNet.

1. table, tabular array -- (*a set of data* ...)
2. table -- (*a piece of furniture* ...)
3. table -- (*a piece of furniture with tableware* ...)
4. mesa, table -- (*flat tableland* ...)
5. table -- (*a company of people* ...)
6. board, table -- (*food or meals* ...)

The **verb** "table" has 1 sense in WordNet.

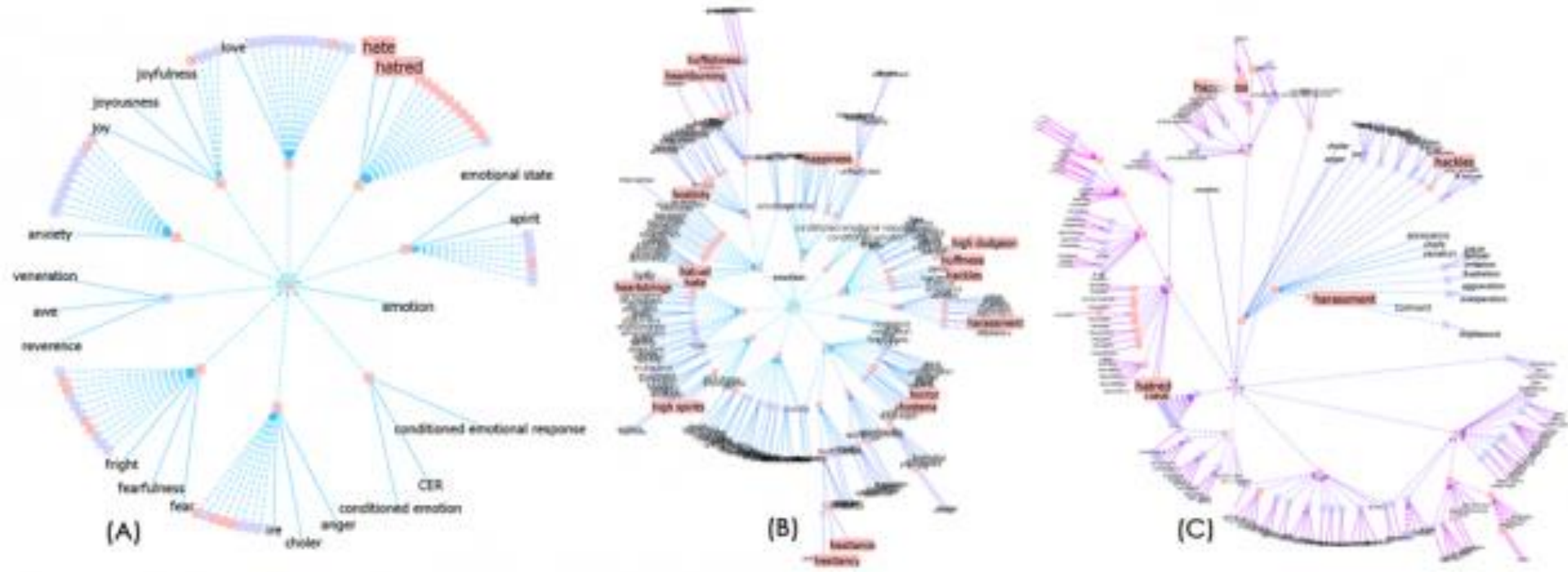
1. postpone, prorogue, hold over, put over, **table**, shelve, set back, defer, remit, put off - (*hold back to a later time; "let's postpone the exam"*)

WordNet Relations (between synsets!)

Relation	Definition	Example
Hypernym	From concepts to superordinates	*breakfast* → *meal*
Hyponym	From concepts to subtypes	*meal* → *lunch*
Has-Member	From groups to their members	*faculty* → *professor*
Member-Of	From members to their groups	*copilot* → *crew*
Has-Part	From wholes to parts	*table* → *leg*
Part-Of	From parts to wholes	*course* → *meal*
Antonym	Opposites	*leader* → *follower*

Relation	Definition	Example
Hypernym	From events to superordinate events	*fly* → *travel*
Troponym	From events to their subtypes	*walk* → *stroll*
Entails	From events to the events they entail	*snore* → *sleep*
Antonym	Opposites	*increase* ⟺ *decrease*

Visualizing Wordnet Relations



C. Collins, "WordNet Explorer: Applying visualization principles to lexical semantics," University of Toronto, Technical Report kmdi 2007-2, 2007.

WordNet Hierarchies: "Vancouver"

WordNet: example from ver1.7.1

For the three senses of "Vancouver"

⇒(city, metropolis, urban center)
⇒ (municipality)
⇒ (urban area)
⇒ (geographical area)
⇒ (region)
⇒ (location)
⇒ (entity, physical thing)

⇒ (administrative district, territorial division)
⇒ (district, territory)
⇒ (region)
⇒ (location
⇒ (entity, physical thing)

⇒ (port)
⇒ (geographic point)
⇒ (point)
⇒ (location)
⇒ (entity, physical thing)

Web interface & API

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: bass [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **bass** (the lowest part of the musical range)
- S: (n) **bass**, bass part (the lowest part in polyphonic music)
- S: (n) **bass**, basso (an adult male singer with the lowest voice)
  - ***direct hypernym*** / *inherited hypernym* / *sister term*
    - S: (n) singer, vocalist, vocalizer, vocaliser (a person who sings)
- S: (n) sea bass, **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- S: (n) freshwater bass, **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- S: (n) **bass**, bass voice, basso (the lowest adult male singing voice)
- S: (n) **bass** (the member with the lowest range of a family of musical instruments)
- S: (n) **bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

- S: (adj) **bass**, deep (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

Wordnet: NLP Tasks

- First success in "obscure" task for Probabilistic Parsing (PP–attachments): words + word–classes extracted from the hypernym hierarchy increase accuracy from 84% to 88% [Stetina and Nagao, 1997]
- Word sense disambiguation
- Lexical Chains (summarization)
- …… and *many others* !

More importantly starting point for larger Ontologies!

More ideas from NLP….

Relations among words and their meanings (*paradigmatic*)

Internal structure of individual words (*syntagmatic*)

Predicate–Argument Structure

- **Represent relationships among concepts, events and their participants**

"I ate a turkey sandwich for lunch"

$\exists$ *w: Isa(w,Eating)* $\wedge$ *Eater(w,Speaker)* $\wedge$
Eaten(w,TurkeySandwich) $\wedge$ *MealEaten(w,Lunch)*

"Nam does not serve meat"

$\exists$ *w: Isa(w,Serving)* $\wedge$ *Server(w, Nam)* $\wedge$
$\neg$*Served(w,Meat)*

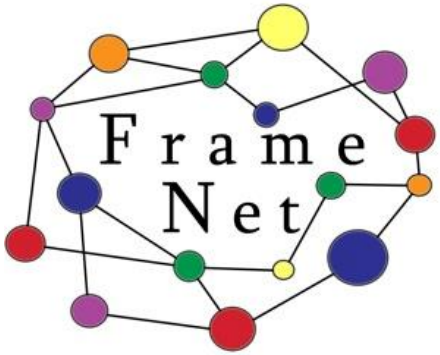
Semantic Roles: Resources

- Move beyond inferences about single verbs

  " *IBM hired John as a CEO* "

  " *John is the new IBM hire* "

  " *IBM signed John for 2M$*"



- FrameNet: Databases containing frames and their syntactic and semantic argument structures
- (book online Version 1.5-update Sept, 2010)
  - for English (versions for other languages are under development)
- FrameNet Tutorial at NAACL/HLT 2015!

FrameNet Entry

Hiring

- **Definition:** An Employer hires an Employee, promising the Employee a certain Compensation in exchange for the performance of a job. The job may be described either in terms of a Task or a Position in a Field.
- **Inherits From:** **Intentionally affect**
- **Lexical Units:** ***commission.n, commission.v, give job.v, hire.n, hire.v, retain.v, sign.v, take on.v***

FrameNet : Semantic Role Labeling

Some roles..
Employer Employee Task Position

- **np-vpto**
  - In 1979 , singer Nancy Wilson HIRED him to open her nightclub act .
  - ….
- **np-ppas**
  - Castro has swallowed his doubts and HIRED Valenzuela as a cook in his small restaurant .

Lecture Overview

- **Ontologies** – what objects/individuals should we represent? what relations (unary, binary,..)?
- Inspiration from **Natural Language**: WordNet and FrameNet
- Extensions based on Wikipedia and mining the Web & Web search logs (YAGO, ProBase, Freebase,……)
- Domain Specific Ontologies (e.g., Medicine: MeSH, UMLS)

YAGO2: huge semantic knowledge base

Derived from **Wikipedia**, **WordNet** and **GeoNames**. (started in 2007, paper in www conference)

10^6 entities (persons, organizations, cities, etc.)

>120* 10^6 facts about these entities.

- **YAGO accuracy of 95%. has been manually evaluated.**
- **Anchored in *time* and *space*. YAGO attaches a *temporal* dimension and a *spatial* dimension to many of its facts and entities.**

Freebase

- "Collaboratively constructed database."
- Freebase contains tens of millions of topics, thousands of types, and tens of thousands of properties and over a billion of facts
- Automatically extracted from a number of resources including **Wikipedia**, **MusicBrainz**, and **NNDB**
- as well as the **knowledge contributed by the human volunteers.**
- Each Freebase entity is assigned a set of human-readable unique keys, which are assembled of a value and a namespace.
- All was **available for free** through the APIs or to download from weekly data dumps

Fast Changing Landscape.....

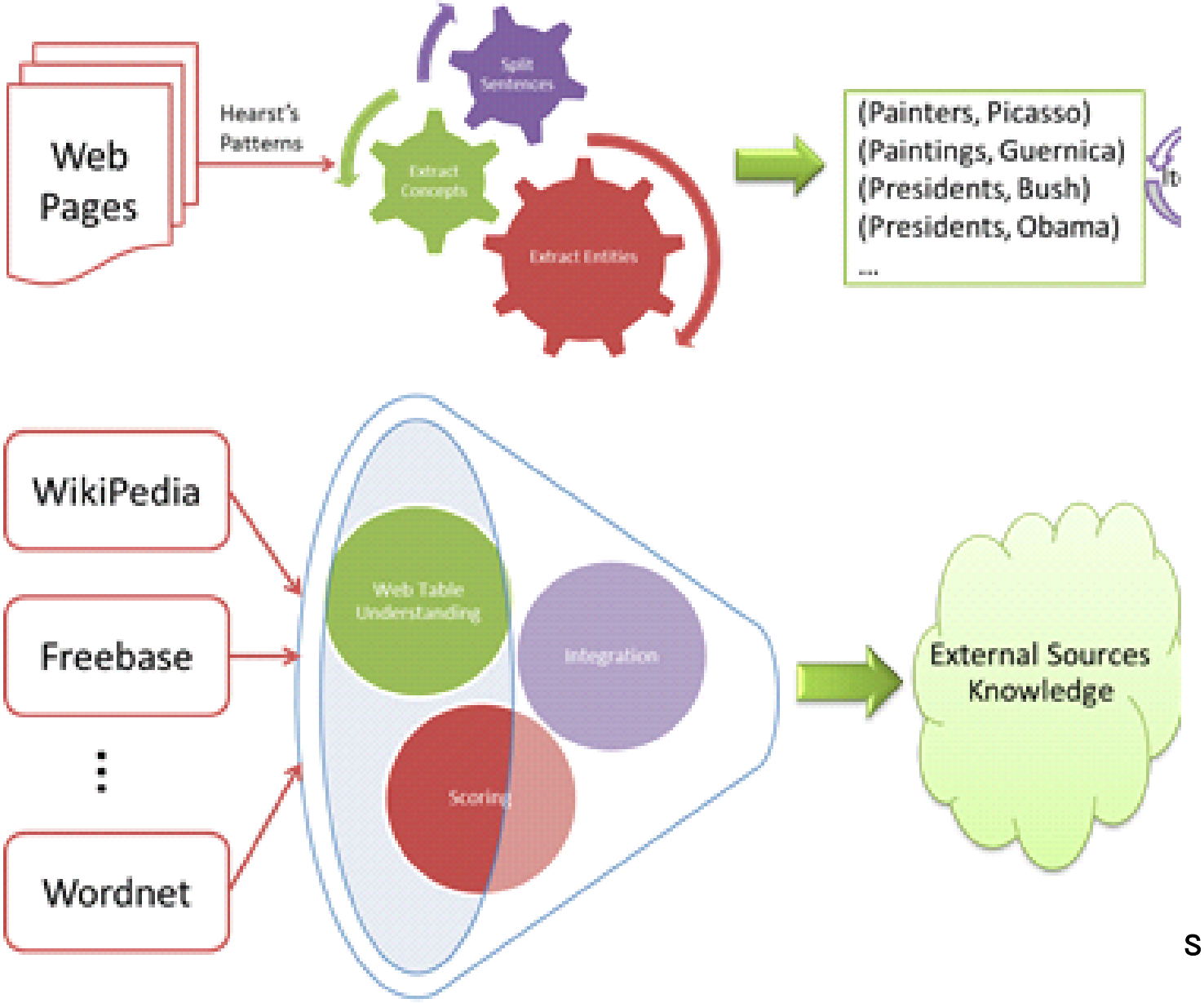
On 16 December 2015, Google officially announced the Knowledge Graph API, which is meant to be a replacement to the Freebase API.

Freebase.com was officially shut down on 2 May 2016.[6]

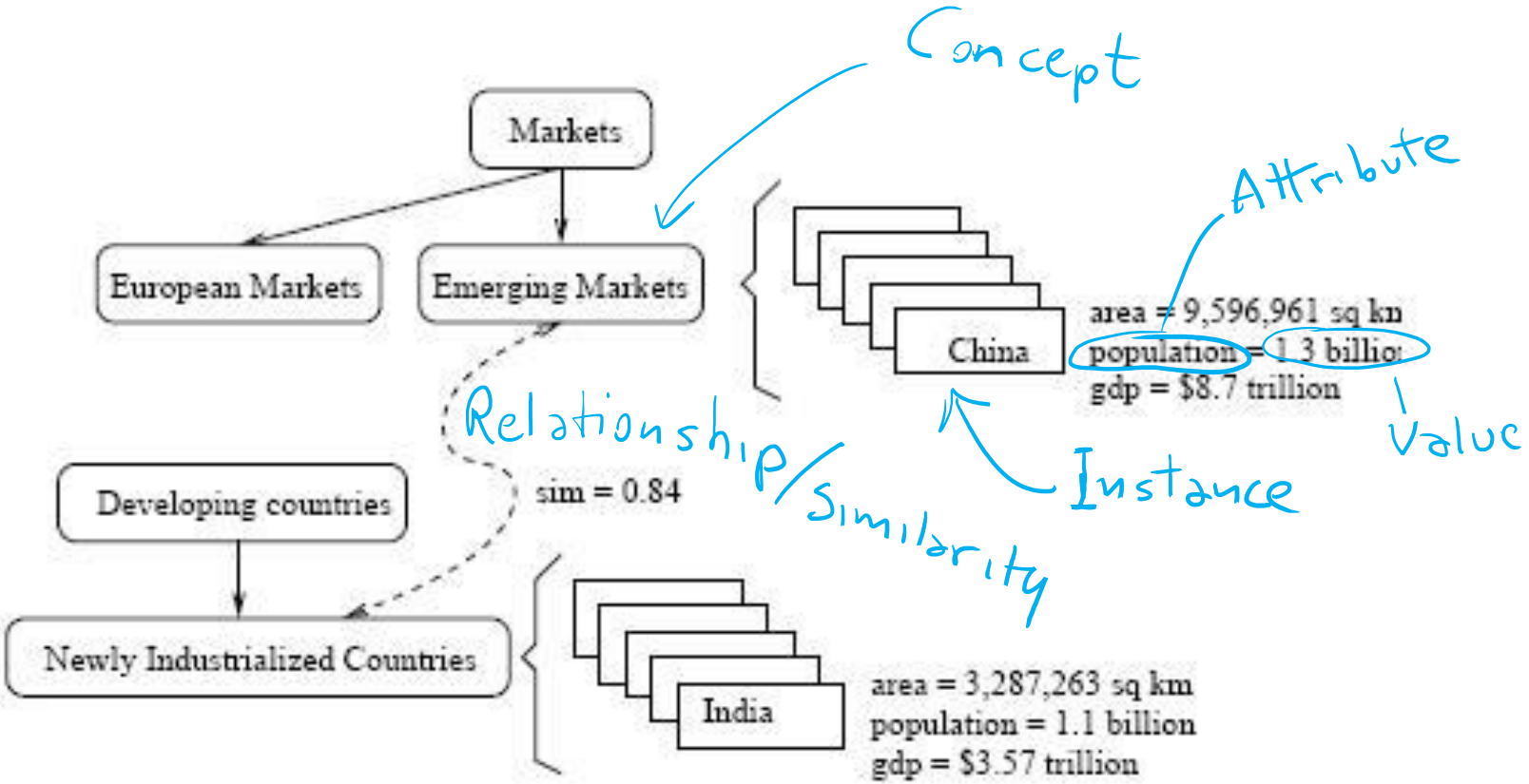
Probase (MS Research) < Sept 2016

- Harnessed from billions of **web pages** and years worth of **search logs**
- Extremely large concept/category space (2.7 million categories).
- Probabilistic model for correctness, typicality (e.g., between concept and instance)

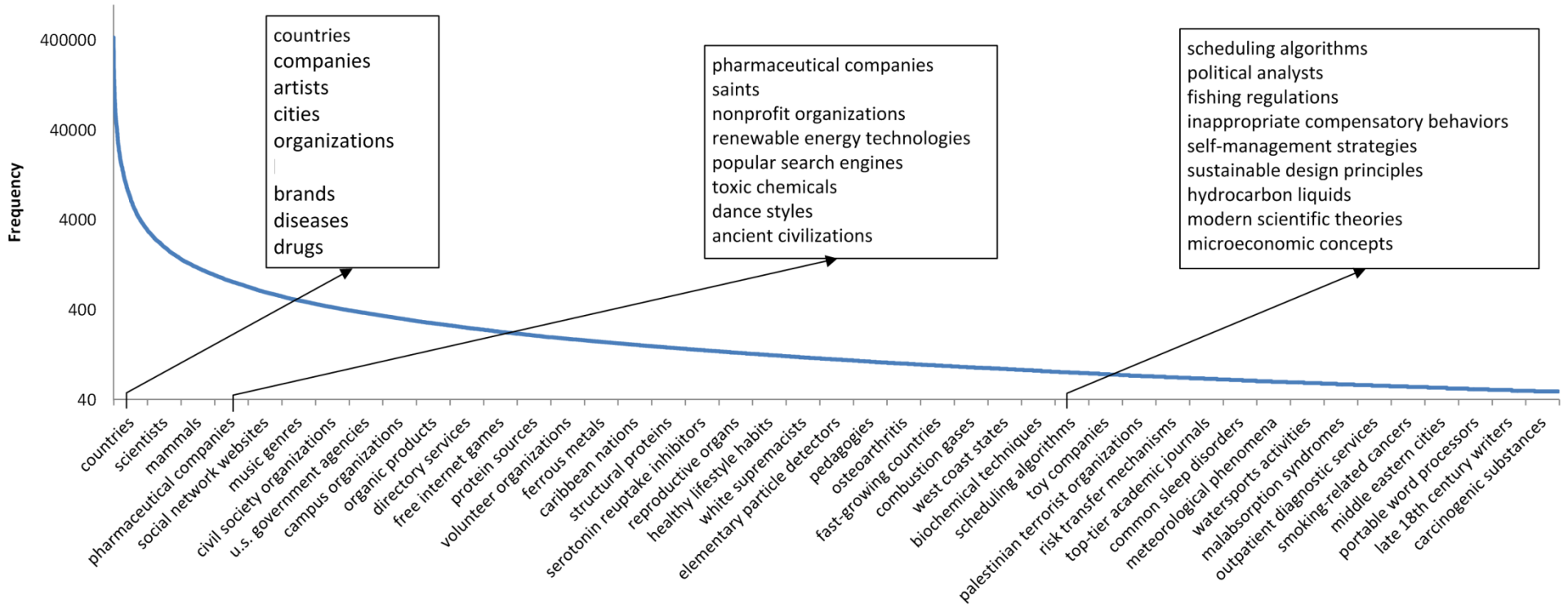
Infrastructure



A snippet of Probase's core taxonomy

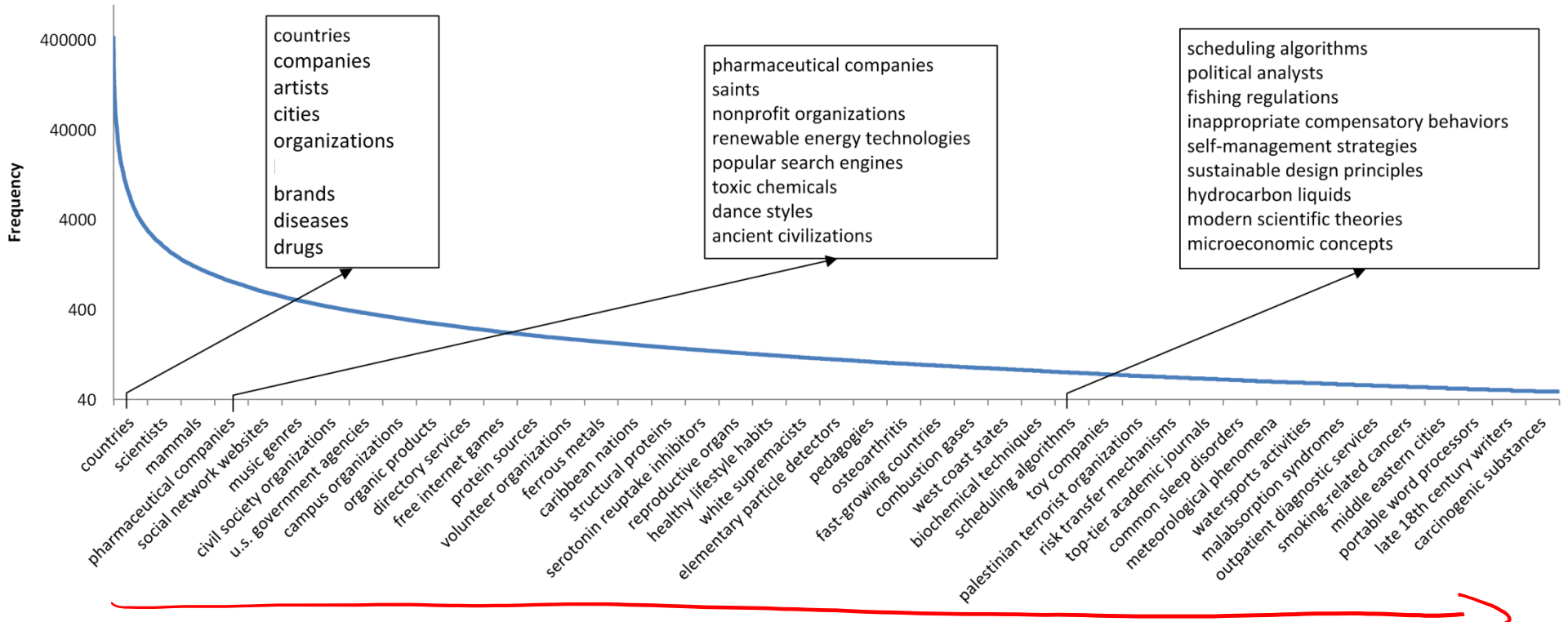


Frequency distribution of the 2.7 million concepts



The Y axis is the number of instances each concept), and on the X axis are the 2.7 million concepts ordered by their size contains(logarithmic scale), and on the X axis are the 2.7 million concepts ordered by their size.

Frequency distribution of the 2.7 million concepts



The Y axis is the number of instances each concept contains(logarithmic scale), and on the X axis are the 2.7 million concepts ordered by their size.

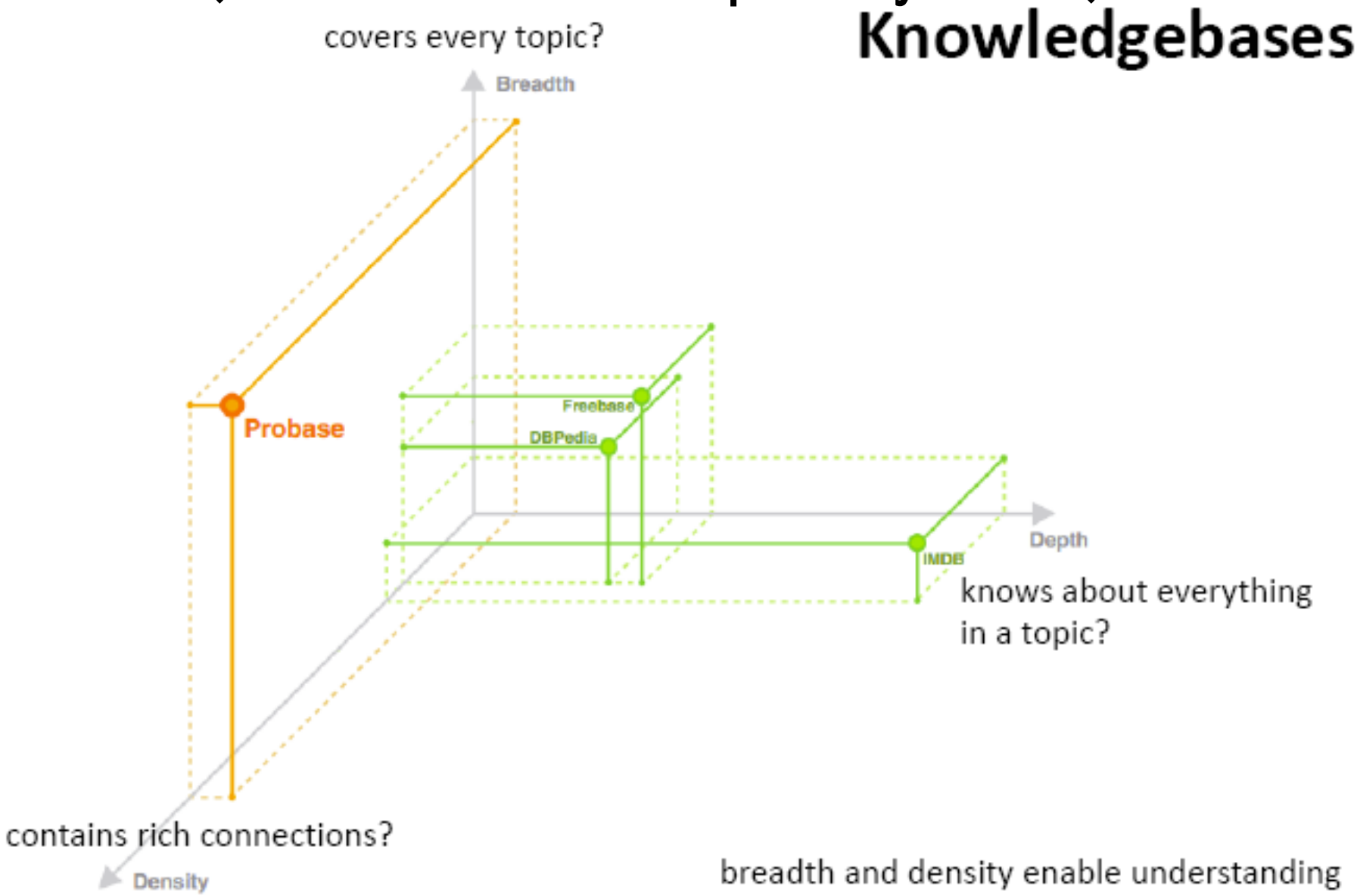
besides popular concepts such as "cities" and "musicians", which are included by almost every general purpose taxonomy, Probase has millions of long tail concepts such as "anti-parkinson treatments", "celebrity wedding dress designers" and "basic watercolor techniques",

Fast Changing Landscape…

From Probase page……

[Sept. **2016**] Please visit our Microsoft Concept Graph release for up-to-date information of this project!

Interesting dimensions to compare Ontologies (but form Probase so possibly biased)



Lecture Overview

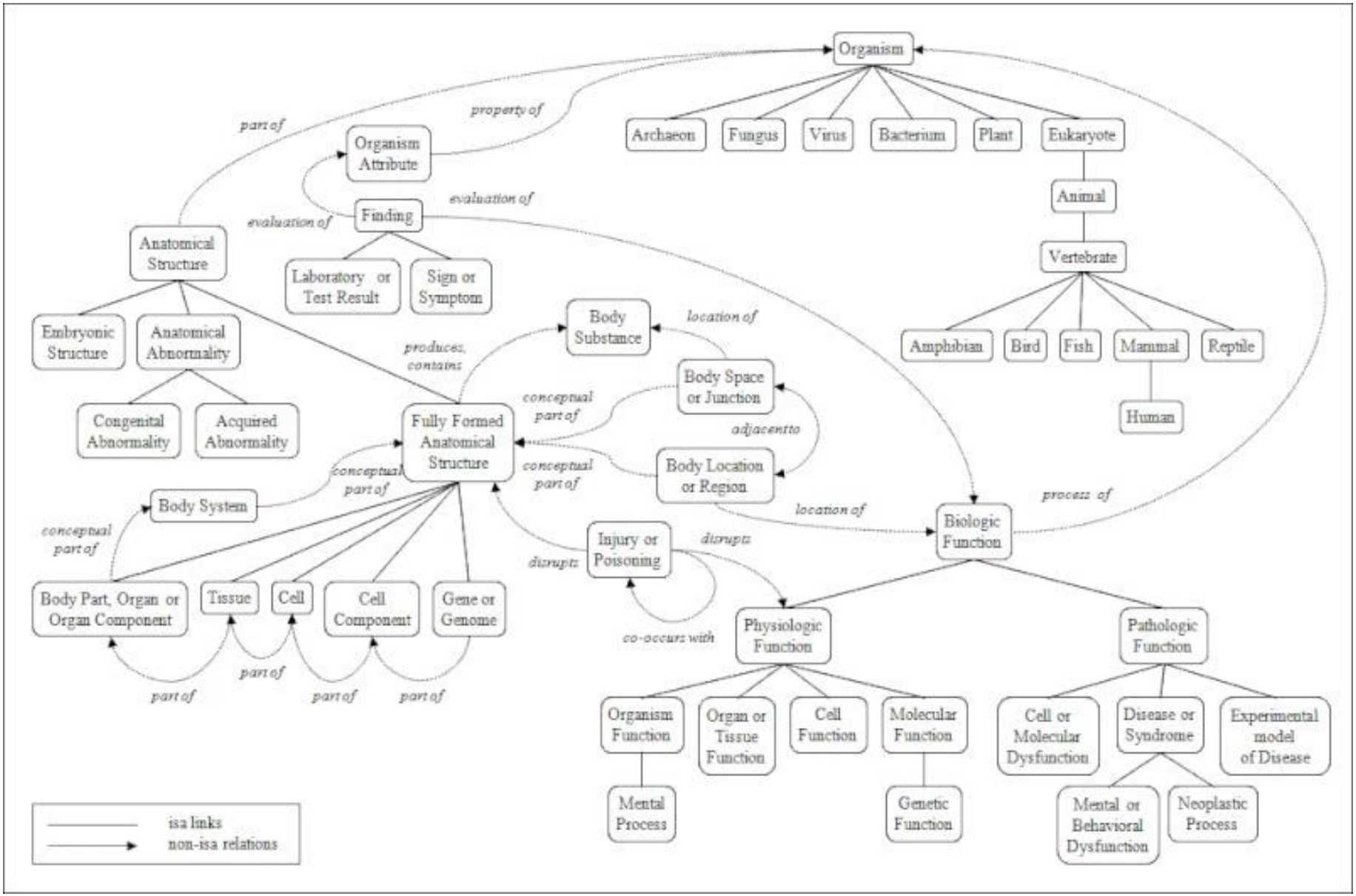
- **Ontologies** – what objects/individuals should we represent? what relations (unary, binary,..)?
- Inspiration from **Natural Language**: WordNet and FrameNet
- Extensions based on Wikipedia and mining the Web (YAGO, ProBase, Freebase)
- Domain Specific Ontologies (e.g., Medicine: MeSH, UMLS)

Domain Specific Ontologies: UMLS, MeSH

Unified Medical Language System: brings together many health and biomedical vocabularies

- Enable interoperability (linking medical terms, drug names)
- Develop electronic health records, classification tools
- Search engines, data mining

Portion of the UMLS Semantic Net



Learning Goals for today's class

You can:

- Define an Ontology
- Describe and Justify the information represented in Wordnet and Framenet
- Describe and Justify the three dimensions for comparing ontologies

Announcements: Midterm

- **Avg 66 Max 103! Min 14**
- **If score below 70 need to very seriously revise all the material covered so far**
- **You can pick up a printout of the solutions along with your midterm**

BUT

Before you look at the solutions **try to answer the questions** by yourself **now that you have all the time you want and access to your notes**

New Re-weighting to help you

Original breakdown

- Assignments -- 15%
- Readings: Questions and Summaries -- 10%
- **Midterm -- 30%**
- **Final -- 45%**

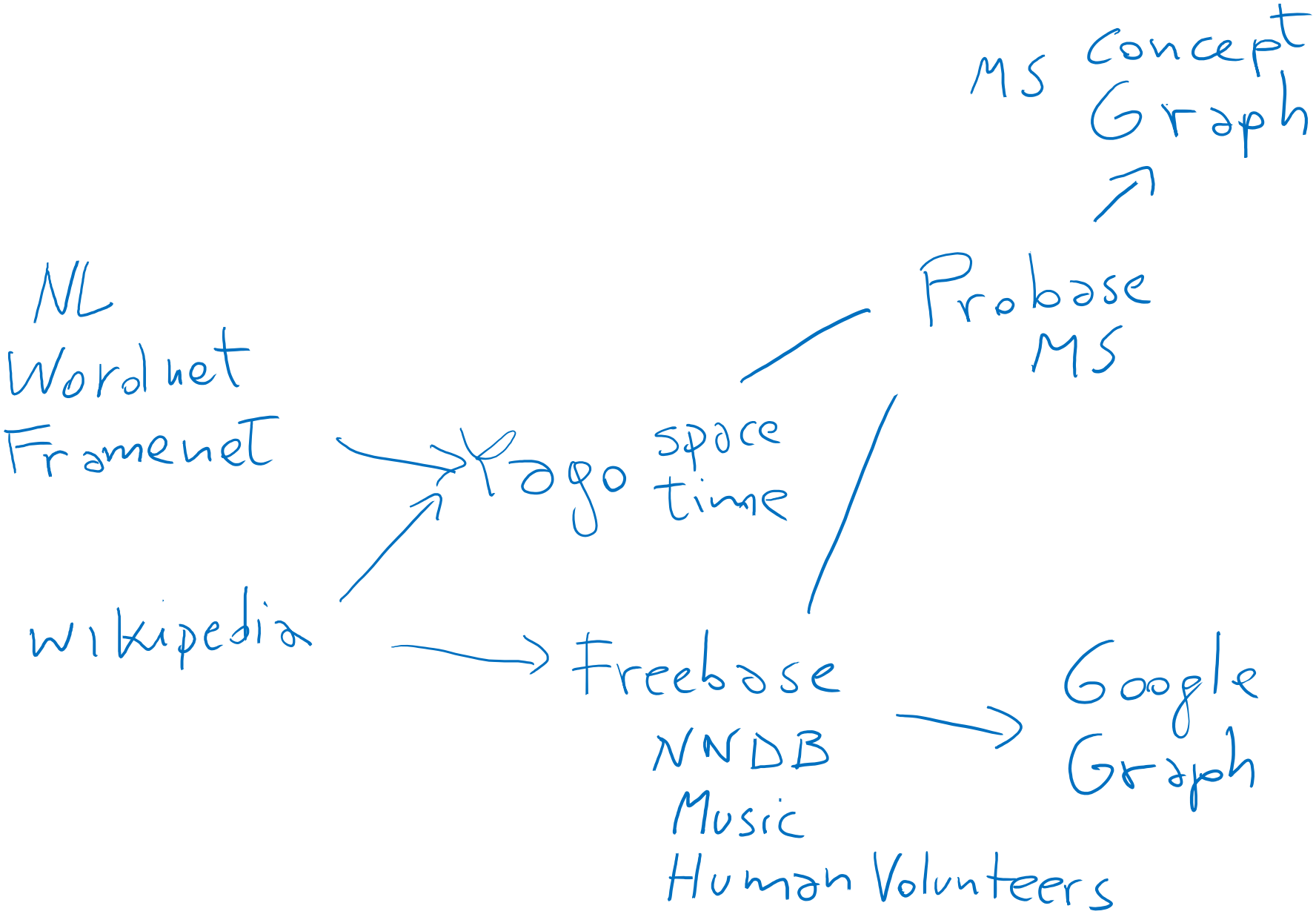
BUT If your grade improves 10% from the midterm to the final

- Assignments -- 15%
- Readings: Questions and Summaries -- 10%
- **Midterm -- 15%**
- **Final -- 60%**

Assignment-3 out – due Nov 20
(8-18 hours – working in pairs on programming parts is strongly advised)

Next class Mon

- **Similarity measures in ontologies (e.g., Wordnet)**


MS Concept Graph
NL
Wordnet
Framenet
Probase
MS
Yago
space
time
wikipedia
Freebase
NNDB
Music
Human Volunteers
Google
Graph

DBpedia is a structured twin ofWikipedia. Currently it describes more than 3.4 million entities. DBpedia resources bear the names of the Wikipedia pages, from which they have been extracted.

YAGO is an automatically created ontology, with taxonomy structure derived from WordNet, and knowledge about individuals extracted from Wikipedia. Therefore, the identifiers of resources describing individuals in YAGO are named as the corresponding Wikipedia pages. YAGO contains knowledge about more than 2 million entities and 20 million facts about them.

Freebase is a collaboratively constructed database. It contains knowledge automatically extracted from a number of resources including Wikipedia, MusicBrainz,2 and NNDB,3 as well as the knowledge contributed by the human volunteers. Freebase describes more than 12 million interconnected entities. Each Freebase entity is assigned a set of human-readable unique keys, which are assembled of a value and a namespace. One of the namespaces is the Wikipedia namespace, in which a value is the name of the Wikipedia page describing an entity.