

MULTILINGUAL WORD SENSE DISAMBIGUATION

"You will know a word by the company it keeps." (Firth)

Submitted By:

Nikita Jain

14535031

Under Supervision Of:

Dr. Dhaval Patel

CS Department

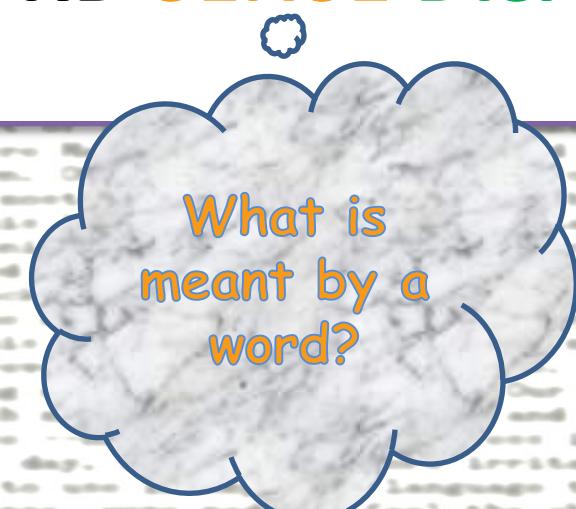
for the most wild, yet most humbly narrative which I am about to pen. I neither expect nor call it belief. Had indeed could I be to expect it, in a case where my very enemies reject their own evidence. Yet, and as I not --and very surely do I not dream. But to-morrow I die, and to-day I would unburden my soul. My immediate purpose is to place before the world, plainly, succinctly, and without comment, a series of mere household events. In their consequences, these events have

MULTILINGUAL WORD SENSE DISAMBIGUATION

What is
meant by a
word?

Ambiguity

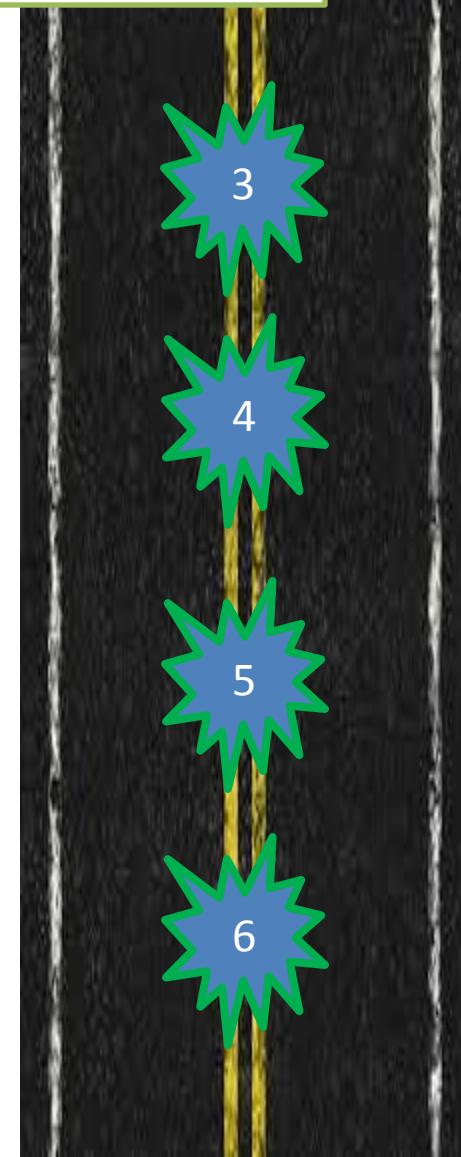
MULTILINGUAL WORD SENSE DISAMBIGUATION



Cross Source Processing will help in enriching knowledge !!

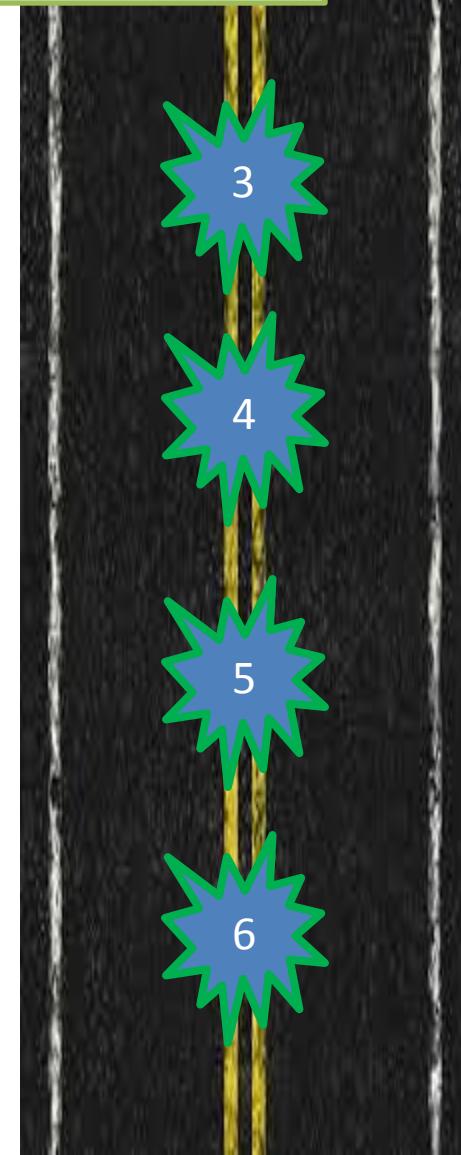
The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



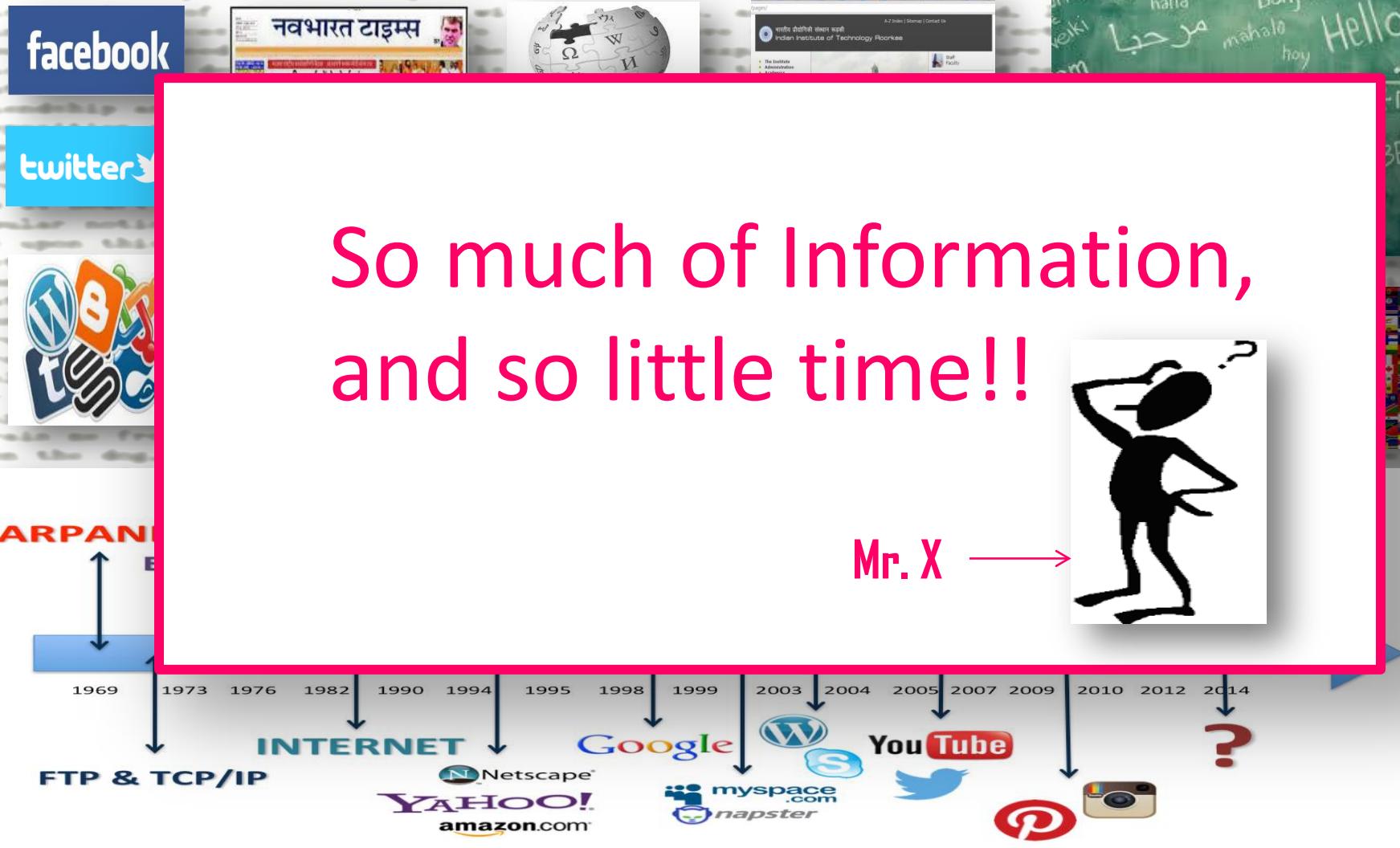
INSPIRATION



WIKIPEDIA
The Free Encyclopedia



INSPIRATION



INSPIRATION



INSPIRATION

facebook

twitter

ARPANET

INTERNET

FTP & TCP/IP

Netscape

YAHOO!

amazon.com

Google

myspace.com

napster

YouTube

Twitter

Pinterest

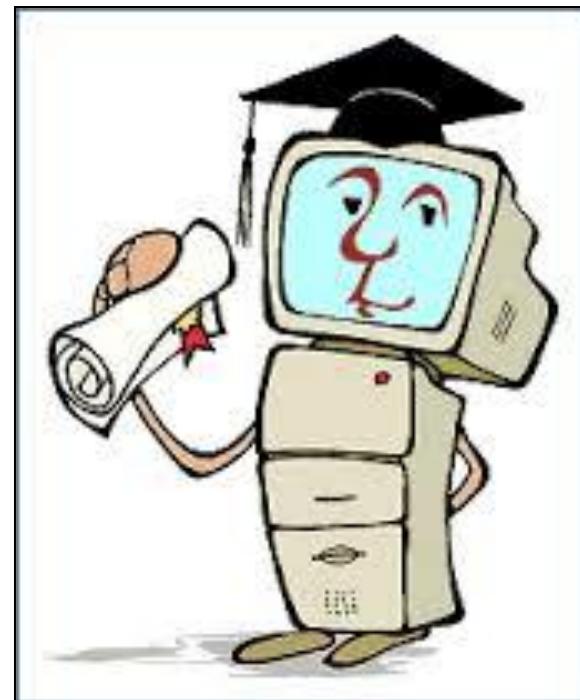
Instagram

?

So what Mr. X can do?

Take help from a computational machine, by making it smarter!

How to make Machine Smart ?



How to make Machine Smart ?

- ✓ The first thing we need to do is,

**Make Machine (process + understand + synthesize)
Natural/Human Language.**

How to make Machine Smart ?

- ✓ Make Machine (process + understand + synthesize) Natural/Human Language.
- ✓ Now, to make above tasks possible, we need



Entity Recognition

How to make Machine Smart ?

- ✓ Make Machine (process + understand + synthesize) Natural/Human Language.
- ✓ Entity Recognition [process]
- ✓ Now, there is a possibility that one mention in a text may match to multiple entities. So, there's a need for
Entity Disambiguation

How to make Machine Smart ?

- ✓ Make Machine (process + understand + synthesize) Natural/Human Language.
- ✓ Entity Recognition [process]
- ✓ Entity Disambiguation [understand]
- ✓ After removing all the ambiguities, we need to link all the mentions to their corresponding entity. For this we need,

Entity Linking

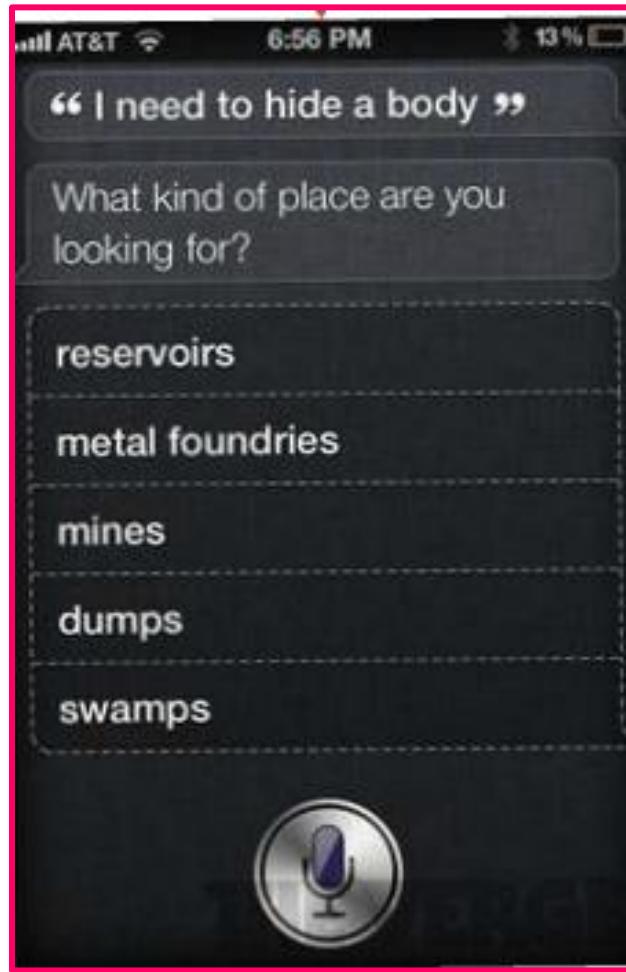
How to make Machine Smart ?

- ✓ Make Machine (process + understand + synthesize) Natural/Human Language.
- ✓ Entity Recognition [process]
- ✓ Entity Disambiguation [understand]
- ✓ Entity Linking [synthesize]
- ✓ Now, it's the time to reap the fruits, with the help of

Entity Oriented Task Extraction



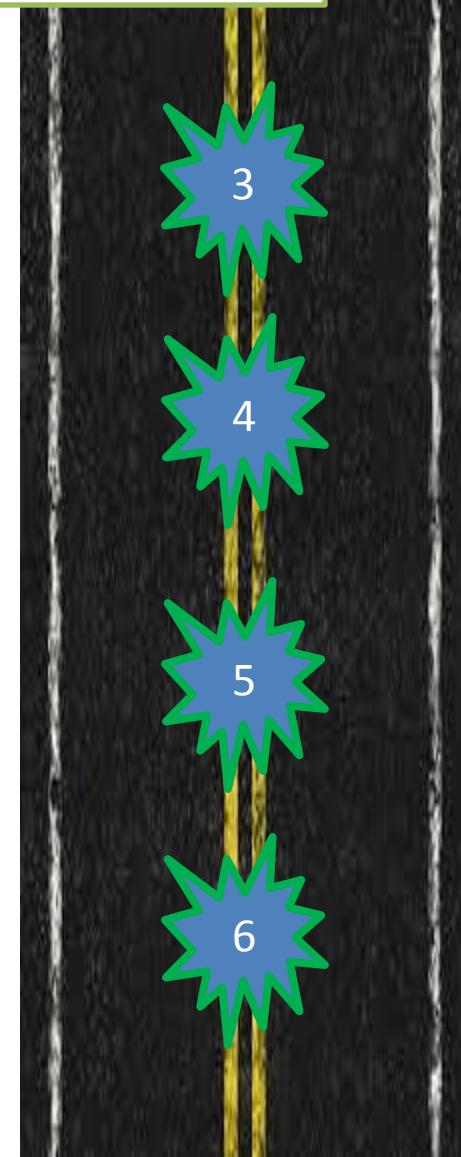
EXAMPLE: Entity Oriented Task Extraction



Siri, Apple Product

The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



Entity Recognition

- ❖ It's a task of recognizing all the entities (such as name, location, organization, thing) mentioned in the text.
- ❖ Example

Roorkee (Hindi: रुड़की Rūṛkī) is a city in Haridwar district, **Uttarakhand**(Hindi:उत्तराखण्ड), that is spread over a flat terrain with the grand spectacle of Himalayas ranges flanking it in the East and the North-east. It is on the banks of the **Ganges canal** on the national highway 58 (Delhi - Sri Badnrinath - Mana) between Delhi and Dehradun. Also the National Highway 73, connecting Roorkee (Uttarakhand) and Panchkula (Haryana) originates from here. The dominant feature of the city is the **Upper Ganges Canal** which flows north-south and bisects the city. Also known for **Roorkee Cantonment**, one of the country's oldest,^[2] and the headquarters of **Bengal Engineer Group** (Bengal Sappers) since 1853.^{[3][4]} The renowned **IIT Roorkee** is located in this city.

[Wikipedia page 'Roorkee'](#)

Entity Recognition in **Idiosyncratic** Document

CHALLENGES

- ❖ Highly Domain Specific
- ❖ New Terms, which may not be observed previously
- ❖ Scarcity of semi-structured informational resources

Entity Recognition in Idiosyncratic Document

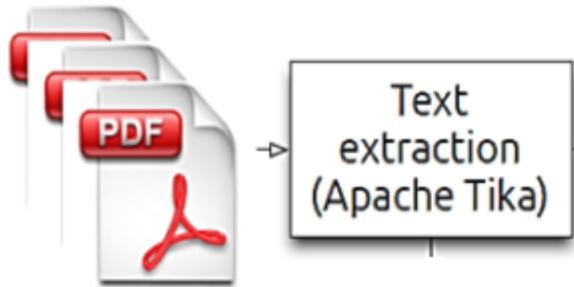
CHAL

Now, What will Mr. X do?

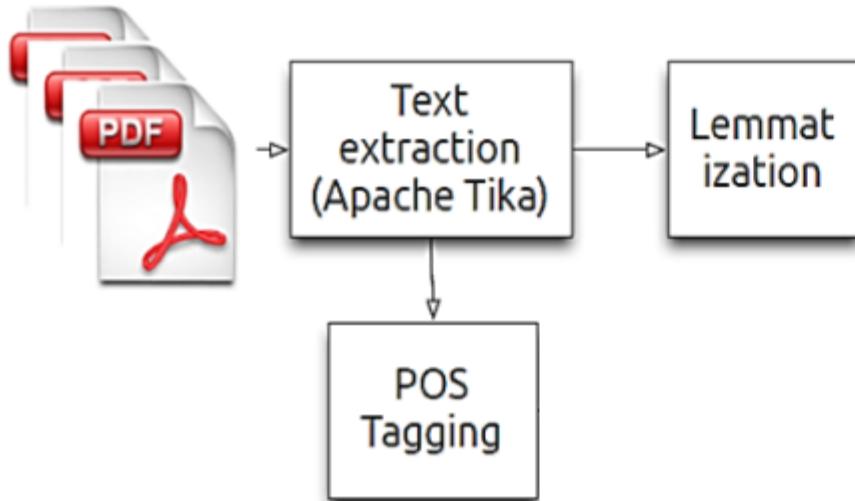
- ❖ Hi
- ❖ Ne
- ❖ So



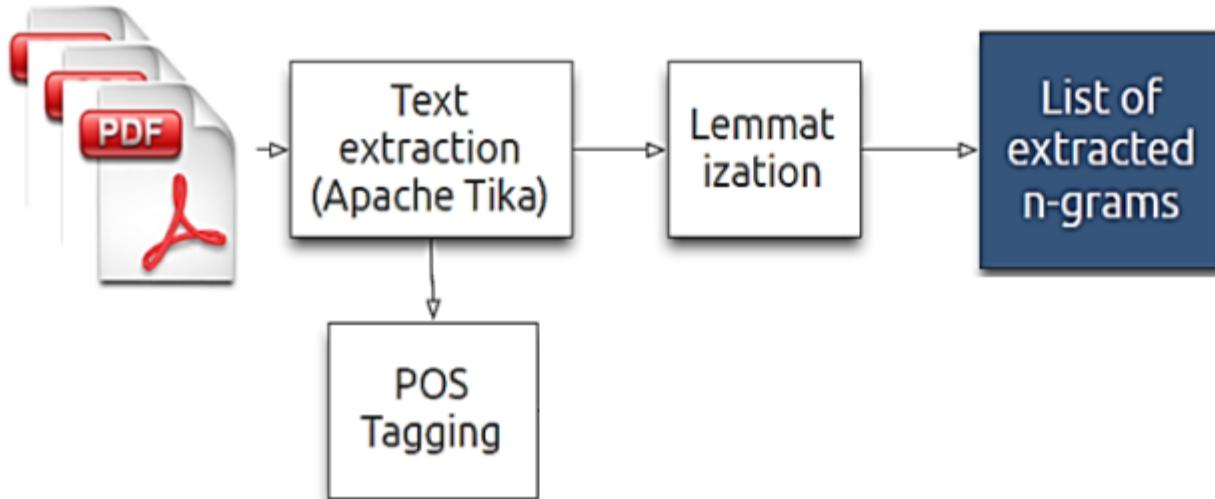
Entity Recognition: Pipeline



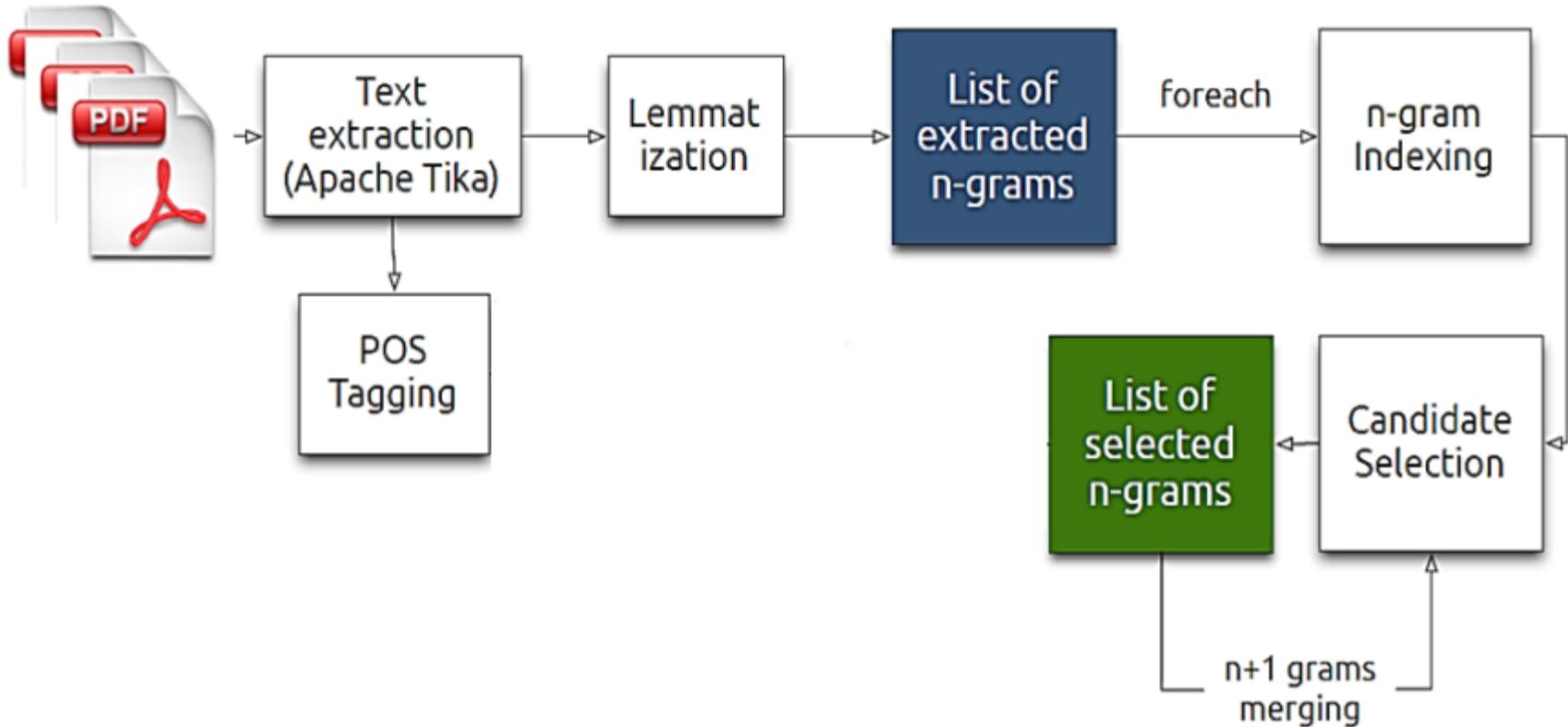
Entity Recognition: Pipeline



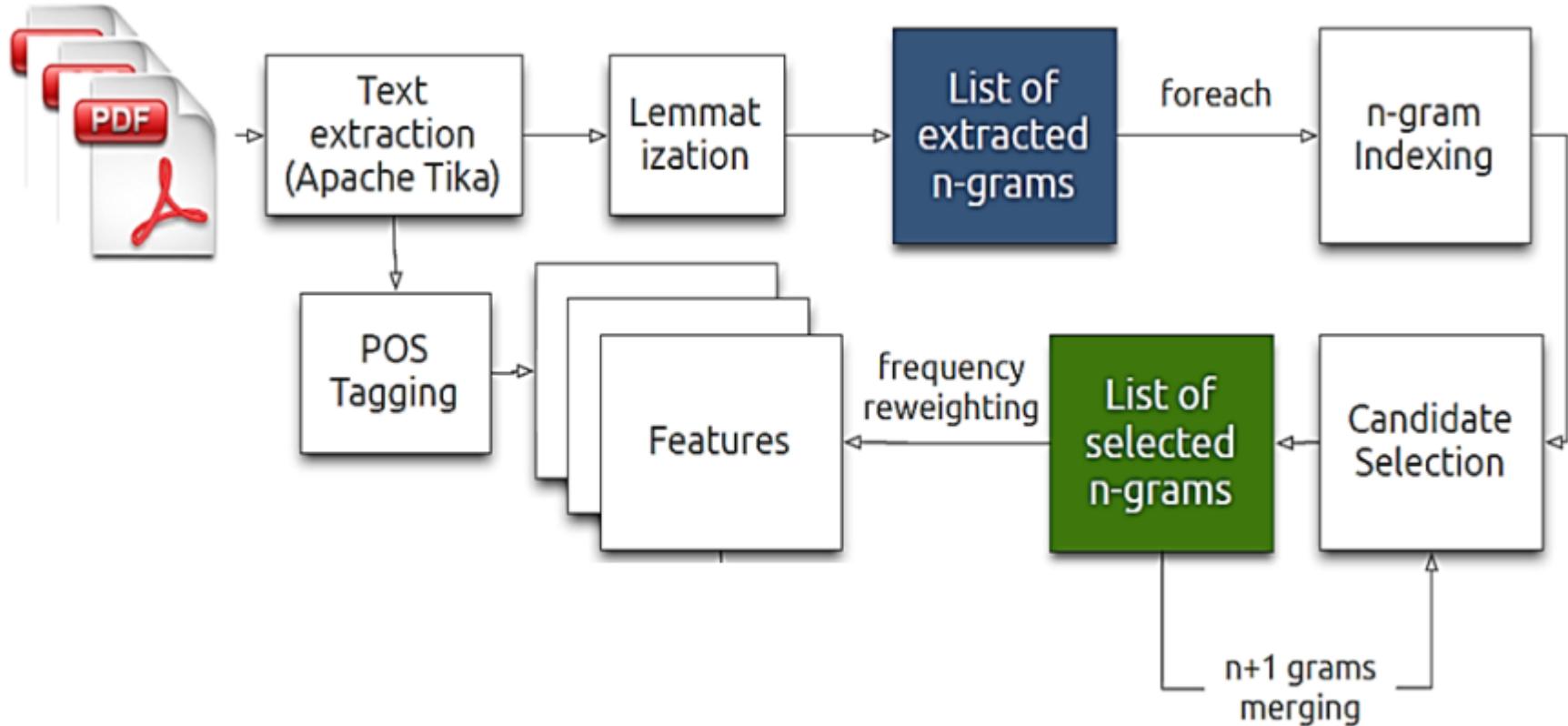
Entity Recognition: Pipeline



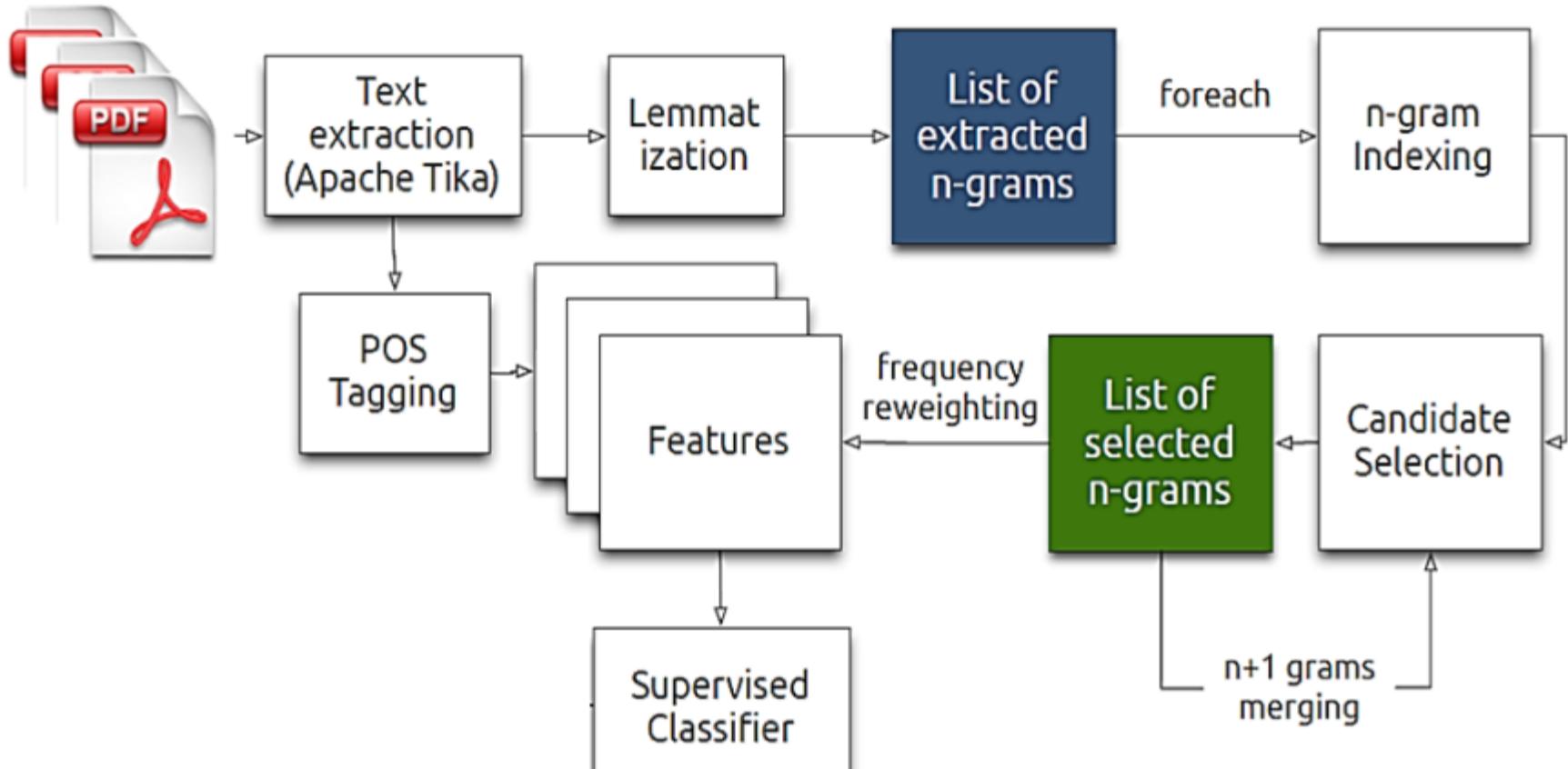
Entity Recognition: Pipeline



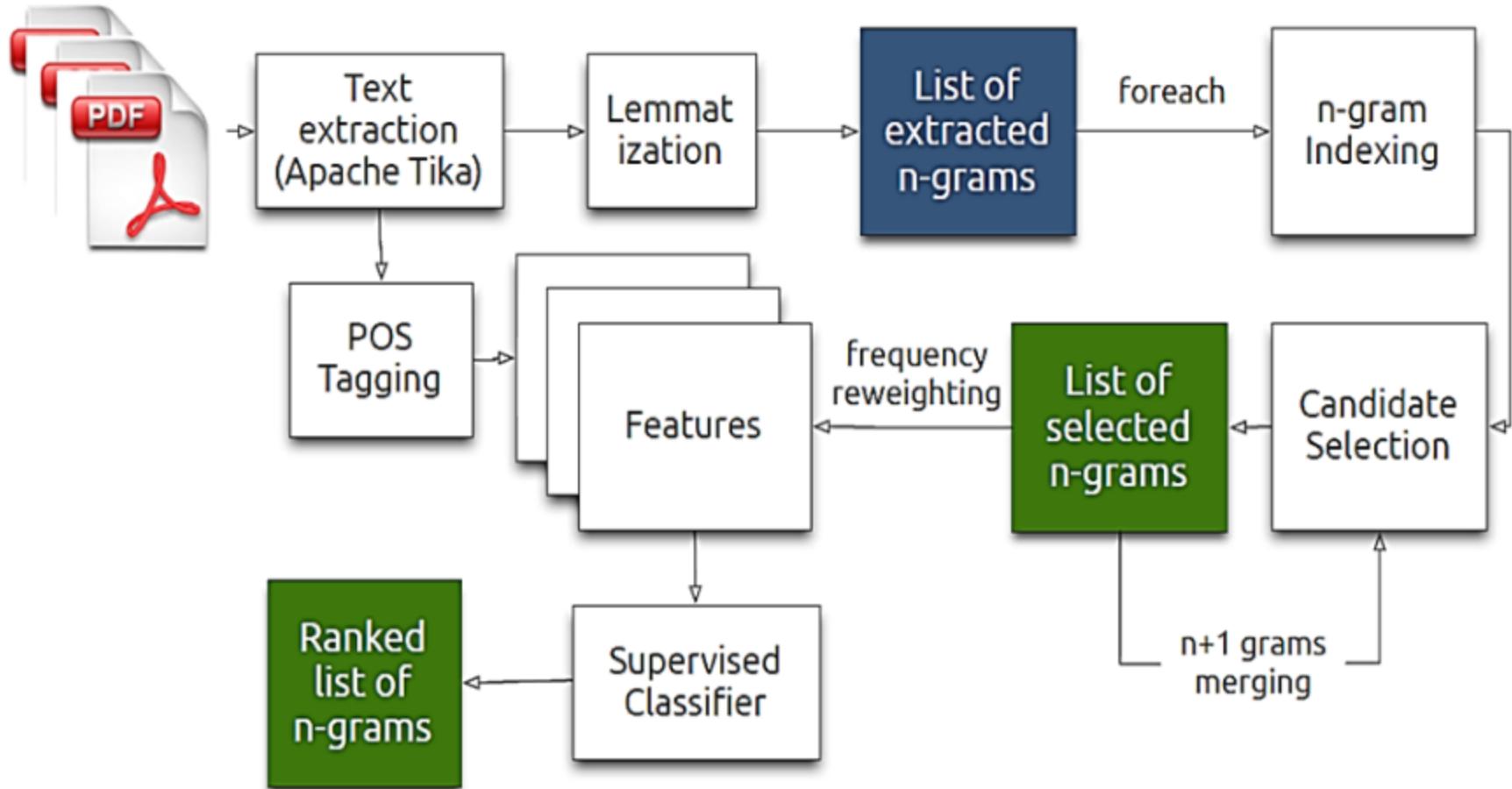
Entity Recognition: Pipeline



Entity Recognition: Pipeline



Entity Recognition: Pipeline



Example

Minerals

Minerals Research Papers explore the ways that minerals can be formed, and their characteristics.

Much like the individual atom is often considered to be the fundamental building block of all matter, so too is the mineral considered to be the fundamental building block of all rocks. According to one author, all rocks are made up of minerals—primarily oxygen and silica. While there are some pure mineral rocks—such as quartz or granite—most rocks are a collection of various minerals combined together to make up one unique substance. While minerals, in the form of rocks are quite common, minerals are unique, in that, they have specific formation properties and structures. To better elucidate these characteristics, this investigation considers the myriad of facets of minerals and their practical use in everyday life.

Despite the fact that minerals are indeed quite common, they can only be formed via three specific mechanisms. First minerals can be formed through the combination of high temperatures—usually in excess of 900°C—and high pressures at the center of the earth. Here, “the earth is composed of a molten silicate mass, called magma, from which most igneous rocks are formed”. As the magma moves toward the surface of the earth—i.e. in a volcanic eruption—it begins to cool and minerals form. In addition to magma, minerals can also be formed when a gas transforms directly into a solid and when water evaporates from a water/mineral solution.

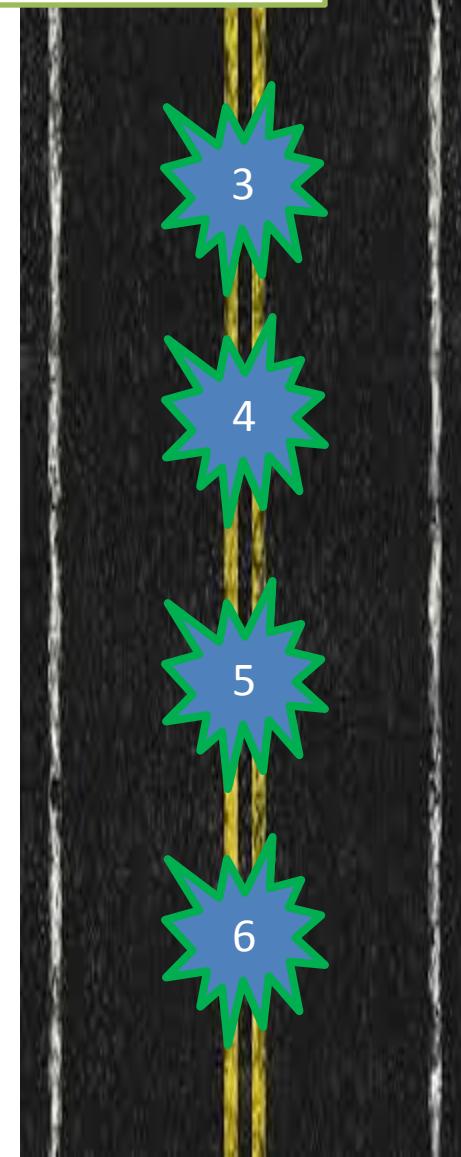


What sets minerals apart from rocks is the fact that minerals consistently have unique geometrical formations that are only characteristic of minerals. What this means is that all of the atoms in a mineral are locked together in a specific crystal lattice that is unique to that mineral. To date scientists have identified six crystal structures that are prevalent among minerals: isometric, tetragonal, hexagonal, orthorhombic, monoclinic and triclinic. While minerals may be similar in composition—i.e. comprised of the same element—their crystal structure is what mitigates their individual chemistry, reactivity and appearance. To illustrate this point, one only needs to consider the carbon minerals of graphite and diamonds. While both are comprised solely of carbon, both have strikingly different properties: graphite is black and soft while diamonds are shiny and hard. The reason for the difference between these two minerals is not the constituent element; rather it is the crystal lattice that comprises each mineral.

Geological Text and Rock highlighted as entity

The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



ENTITY (Word) SENSE DISAMBIGUATION

- ❖ Natural/Human Language is ambiguous.
- ❖ So many words can be interpreted in multiple ways, depending on the context they appear in.

Example:

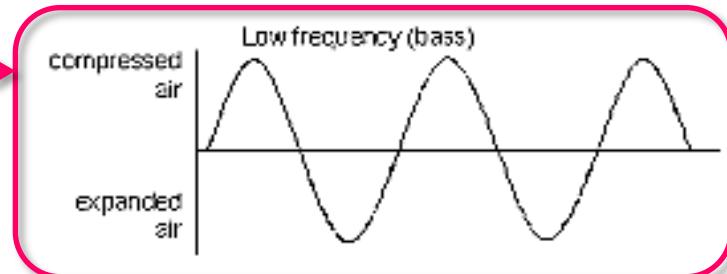
- (1) I can hear *bass* sounds.
- (2) They like grilled *bass*.

ENTITY (Word) SENSE DISAMBIGUATION

- ❖ Natural/Human Language is ambiguous.
- ❖ So many words can be interpreted in multiple ways, depending on the context they appear in.

Example:

- (1) I can hear *bass* sounds.
- (2) They like grilled *bass*.



ENTITY (Word) SENSE DISAMBIGUATION

- ❖ Natural/Human Language is ambiguous.
- ❖ So many words can be interpreted in multiple ways, depending on the context they appear in.

Example:

- (1) I can hear *bass* sounds.
- (2) They like grilled *bass*.



ENTITY (Word) SENSE DISAMBIGUATION

- ❖ Natural/Human Language is ambiguous.
- ❖ So many words can be interpreted in multiple ways, depending on the context they appear in.

Example:

- (1) I can hear *bass* sounds.
- (2) They like grilled *bass*.

- ❖ **WSD:** The association of a given word in a text with a definition or meaning, distinguishable from other possible meanings.

Word Sense Disambiguation: Pipeline

ગુજરાત કે ગાંધીનગર મેં લગે આડવાળી કી ગુમશુદગી કે પોસ્ટર



POS Tagger

Word Sense Disambiguation: Pipeline

ગુજરાત કે ગાંધીનગર મેં લગે આડવાળી કી ગુમશુદગી કે પોસ્ટર



POS Tagger



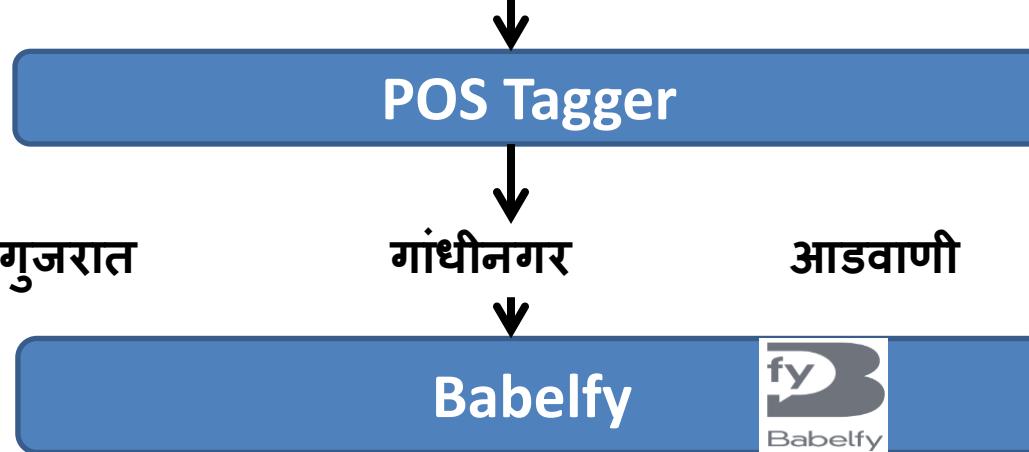
ગુજરાત

ગાંધીનગર

આડવાળી

Word Sense Disambiguation: Pipeline

ગુજરાત કે ગાંધીનગર મેં લગે આડવાળી કી ગુમશુદગી કે પોસ્ટર



Word Sense Disambiguation: Pipeline

ગુજરાત કે ગાંધીનગર મેં લગે આડવાળી કી ગુમશુદગી કે પોસ્ટર



POS Tagger

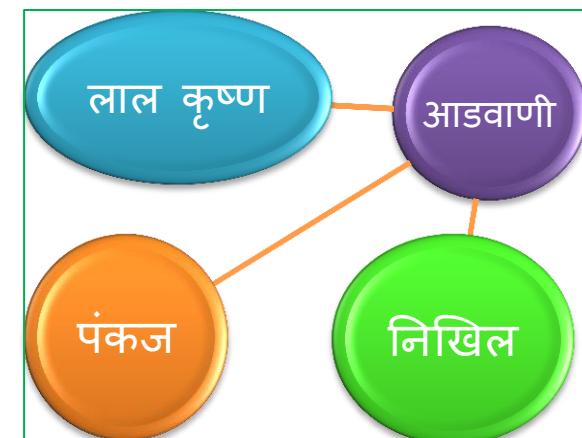
ગુજરાત

ગાંધીનગર

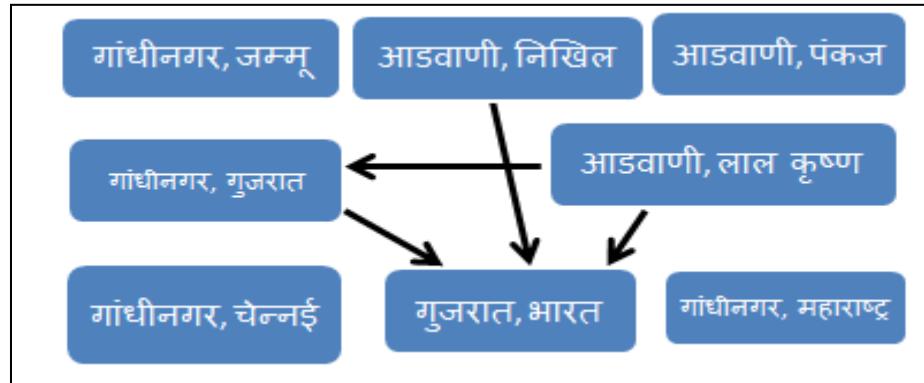
આડવાળી



Babelfy

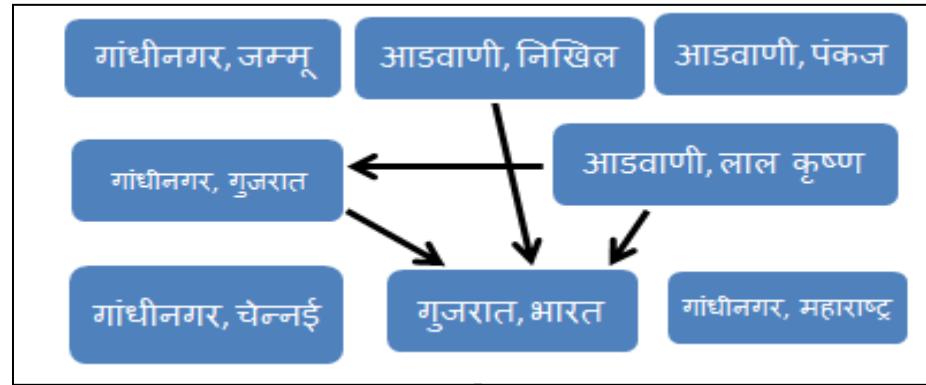


Word Sense Disambiguation: Pipeline

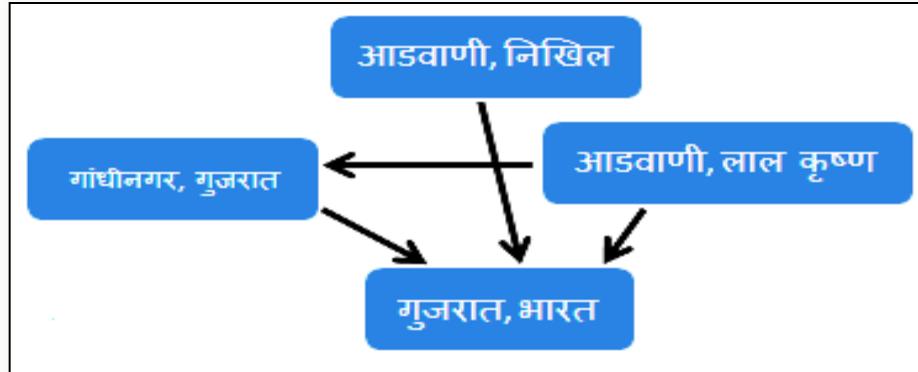


Dense Heuristic Graph Algorithm

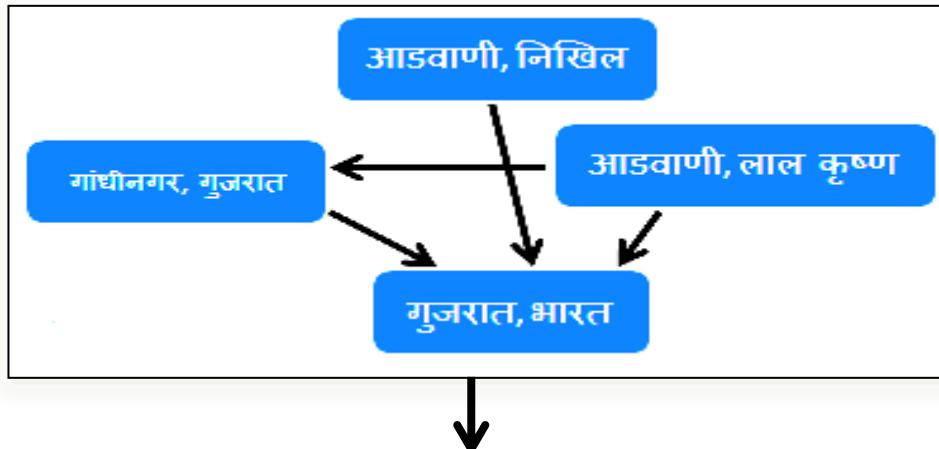
Word Sense Disambiguation: Pipeline



Dense Heuristic Graph Algorithm

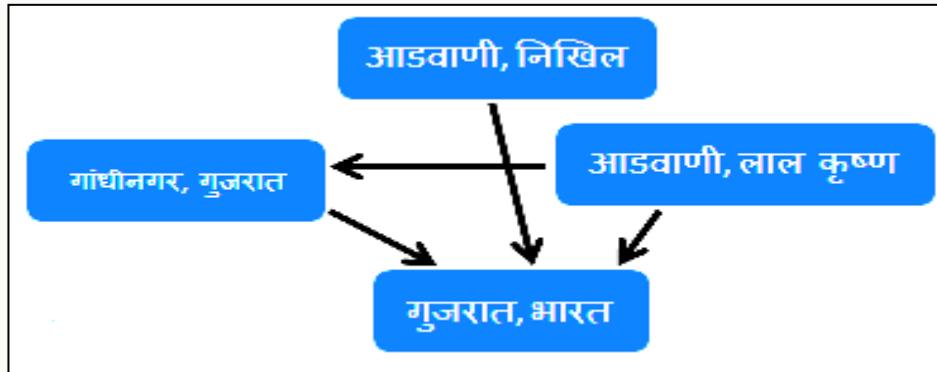


Word Sense Disambiguation: Pipeline



Mapping Entity with the Highest Ranking Candidate

Word Sense Disambiguation: Pipeline



Mapping Entity with the Highest Ranking Candidate

The screenshot shows a search interface with the query "ગુજરાત કે ગાંધીનગર મેં લગે આડવાણી કી ગુમશુદગી કે પોસ્ટર". Below the query, there are three cards with images and descriptions:

- ગુજરાત**: An image of the state emblem of Gujarat. Description: "a region of western India to the north of Bombay (bordering the Arabian Sea)...."
- ગાંધીનગર**: An image of the Gandhinagar city skyline. Description: "Gandhinagar is the capital of the state of Gujarat in western India."
- આડવાણી**: An image of Lal Krishna Advani. Description: "Lal Krishna Advani known as L.K Advani is a Veteran Indian politician."

Verb: Word Sense Disambiguation

- ❖ Up till now our focus was on noun phrases.
- ❖ For understanding of text, know the relations among the entities.
- ❖ For this, we need to disambiguate verbs and verbal phrases also, as verbs are the central element in a sentence.

Verb: Word Sense Disambiguation

Q. What is the balance of the project ?



Weighing Scale



Money

Verb: Word Sense Disambiguation

Q. What is the balance of the project ?



Weighing Scale



Money



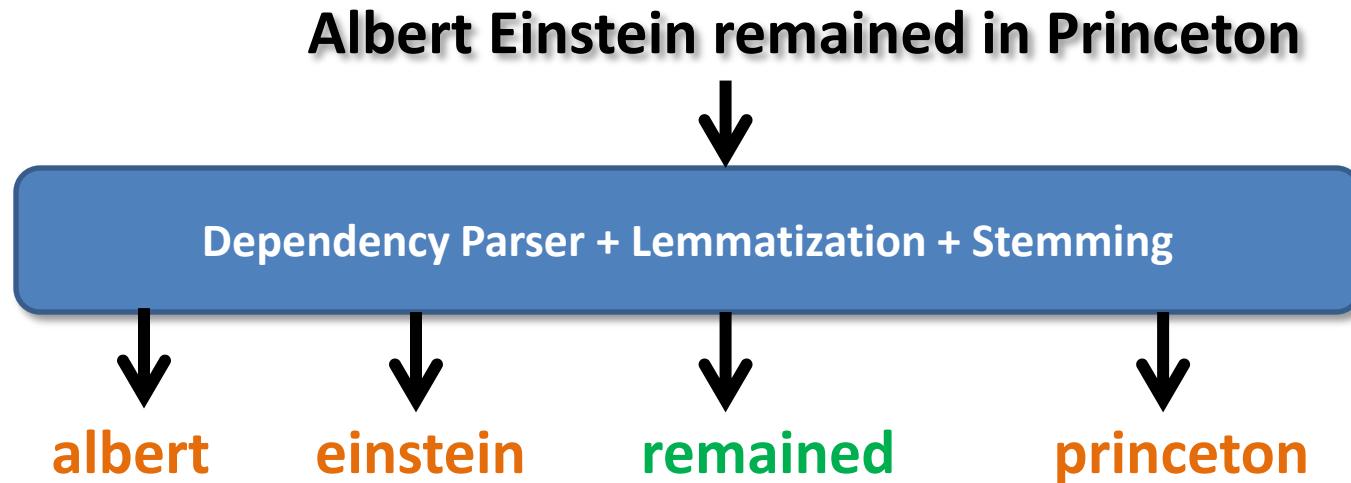
Verb: Word Sense Disambiguation

Albert Einstein remained in Princeton



Dependency Parser + Lemmatization + Stemming

Verb: Word Sense Disambiguation



(*)

For simplicity showing the procedure only for **Verbs**

Verb: Word Sense Disambiguation

Albert Einstein remained in Princeton

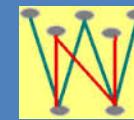


Dependency Parser + Lemmatization + Stemming

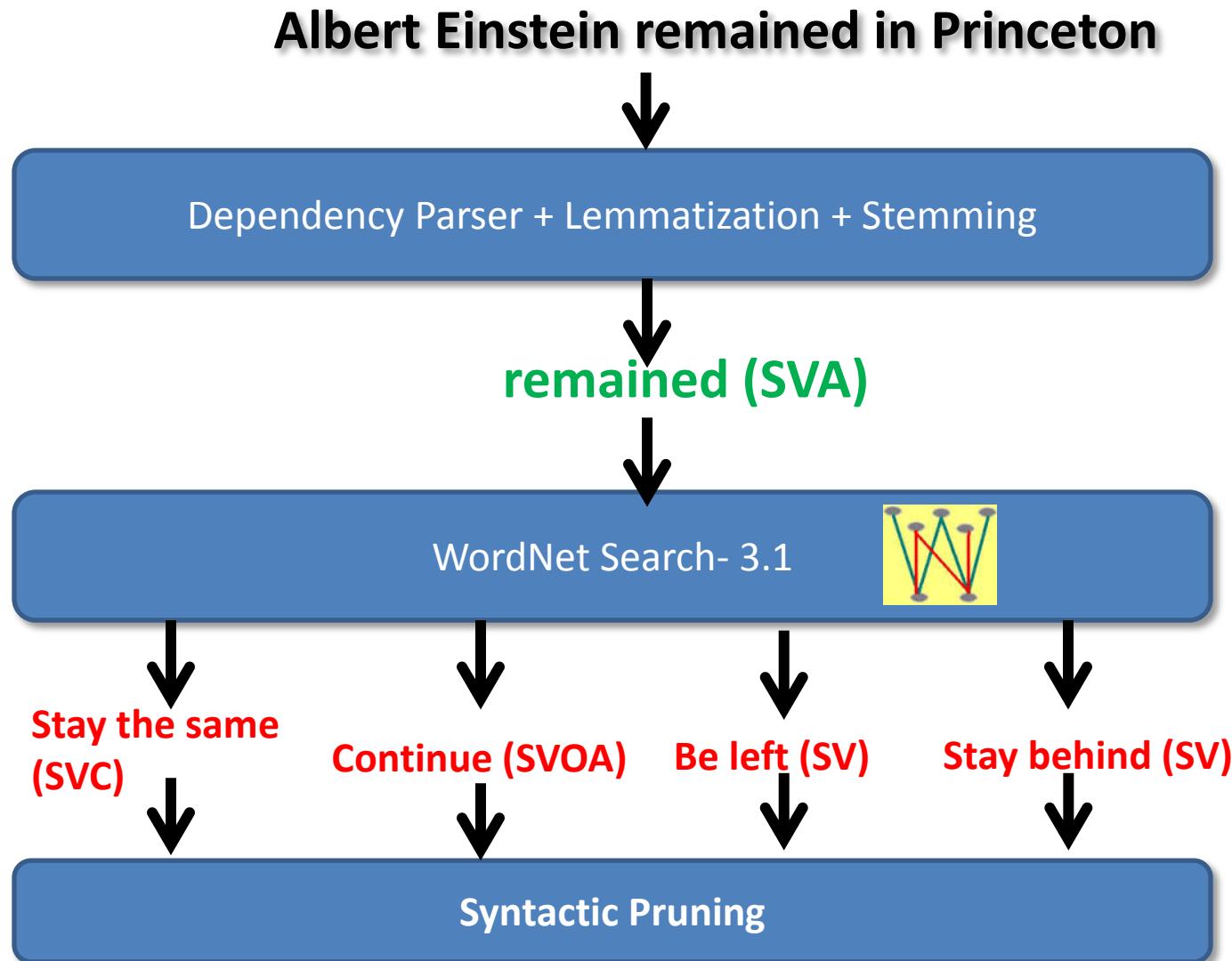


remained (SVA)

WordNet Search- 3.1



Verb: Word Sense Disambiguation



Verb: Word Sense Disambiguation

Albert Einstein remained in Princeton



Dependency Parser + Lemmatization + Stemming + WordNet
Search- 3.1 

Stay the same
(SVC)

Continue(SVOA)

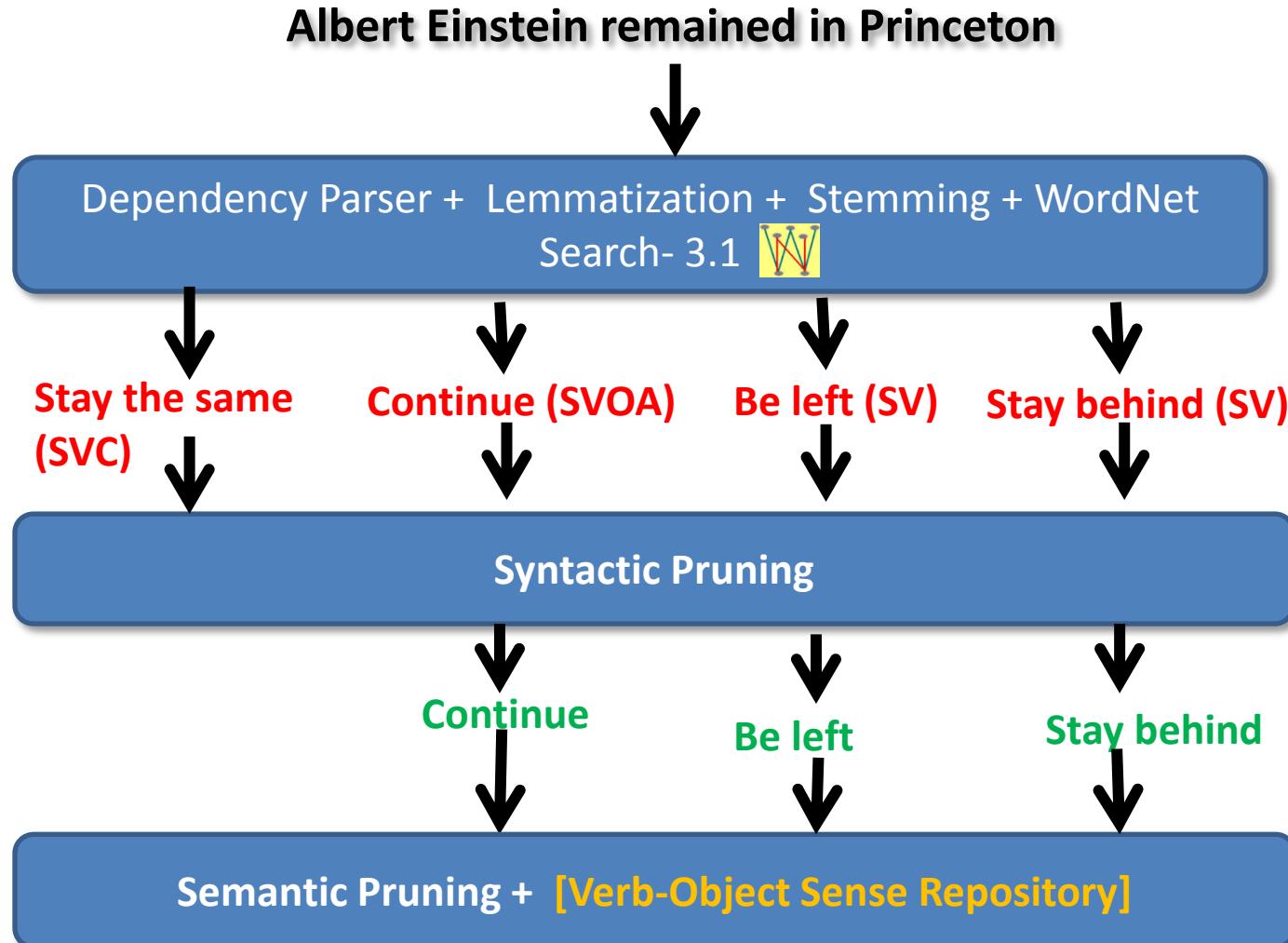
Be left (SV)

Stay behind (SV)

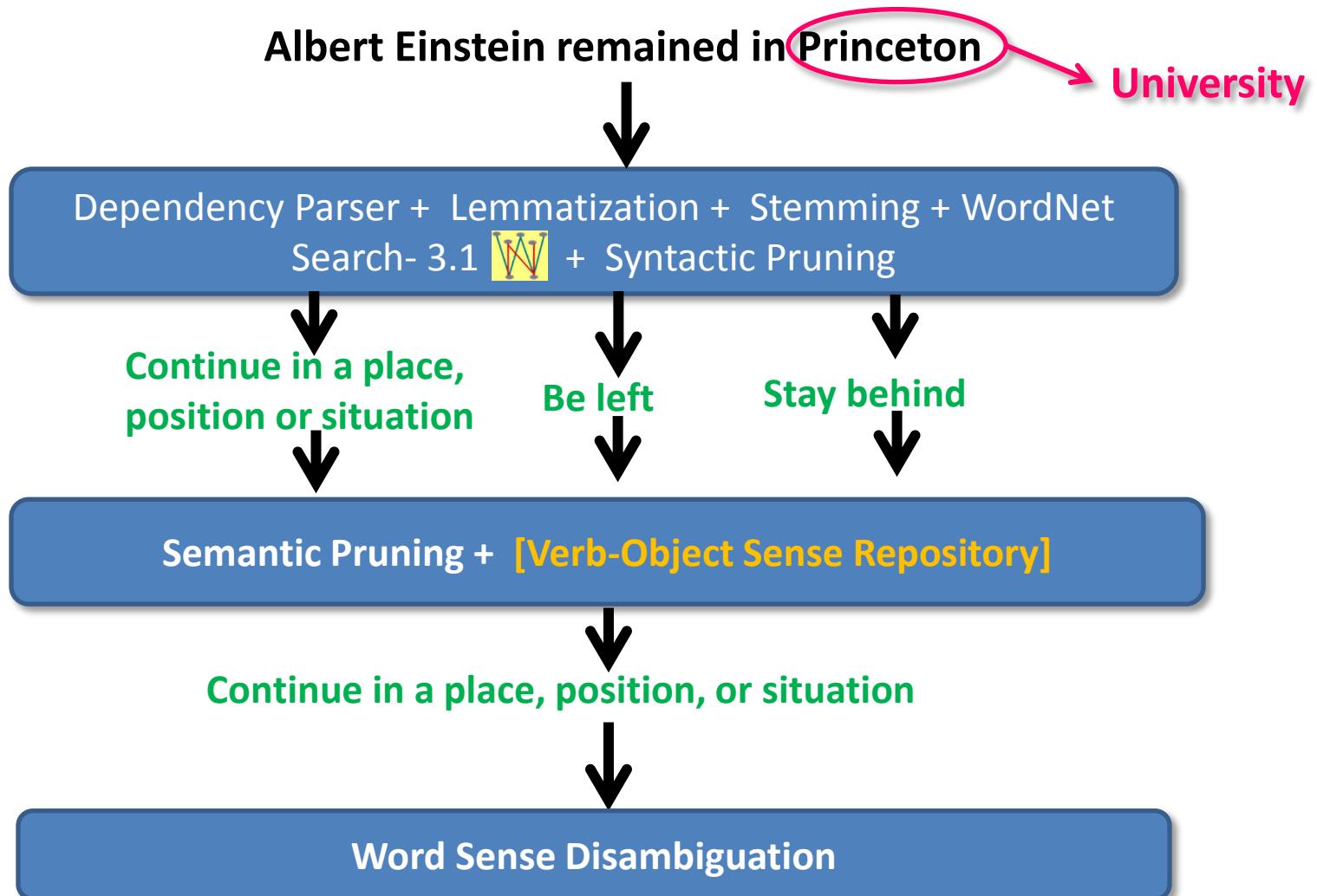


Syntactic Pruning

Verb: Word Sense Disambiguation

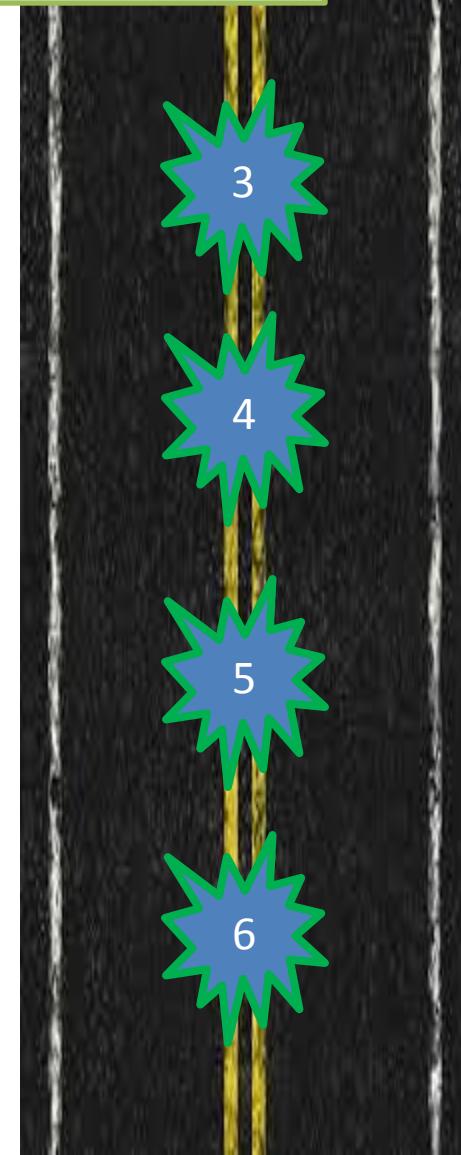


Verb: Word Sense Disambiguation



The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



Entity Linking

- ❖ It's a record linkage or entity resolution,
- ❖ Aligns a textual mention of a entity to an appropriate entry in a knowledge base, which may or may not contain the entity.
- ❖ One of the famous task based on this is '**Wikification**'

Roorkee is well connected to [Delhi](#) by rail and road. It is situated on [National Highways 58 and 73](#) and is on Amritsar-Howrah main rail route.

Some trains which are convenient f [Delhi Shatabdi Express](#) and [Dehradun Janshatabdi Express](#). The nearest airport is located at Jollygrant, Dehradun. Ho

Categories: city district

National Highway 58 from New Delhi to the higher reaches of the therefore, well connected to all the major cities in the region by place from the Maharana Pratap Inter State Bus Terminal (Kash)

Roorkee
From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding reliable sources. Unsourced material may be challenged and removed.

National Highway (India)
From Wikipedia, the free encyclopedia
(Redirected from Indian highways)

This article is about national highways. For expressways network (in Hindi: राष्ट्रीय रा...
रवार्य of India. These highway

Entity Linking: State-of-the-Art

- ❖ **Prior Popularity:** Link a mention to most probable underlying concept
- ❖ **Context Similarity:** Link a mention to most overlapping context
- ❖ **Topical Coherence:** A mention's referent concept is topically coherent with other mention's with the same context.
- ❖ **Colocation and Document Similarity:** The search based on context



Entity Linking

Indian Institute of Technology, formerly the University of Roorkee (1948–2001) and the Thomason College of Civil Engineering (1853–1948), is a public university located in Roorkee, Uttarakhand, India. Established in 1847 in British India by the then lieutenant governor, Sir James Thomason. The Institute admits students to B. Tech., B. Arch. and integrated M.Tech/M.Sc courses through JEE and GATE conducted at centers all over India.



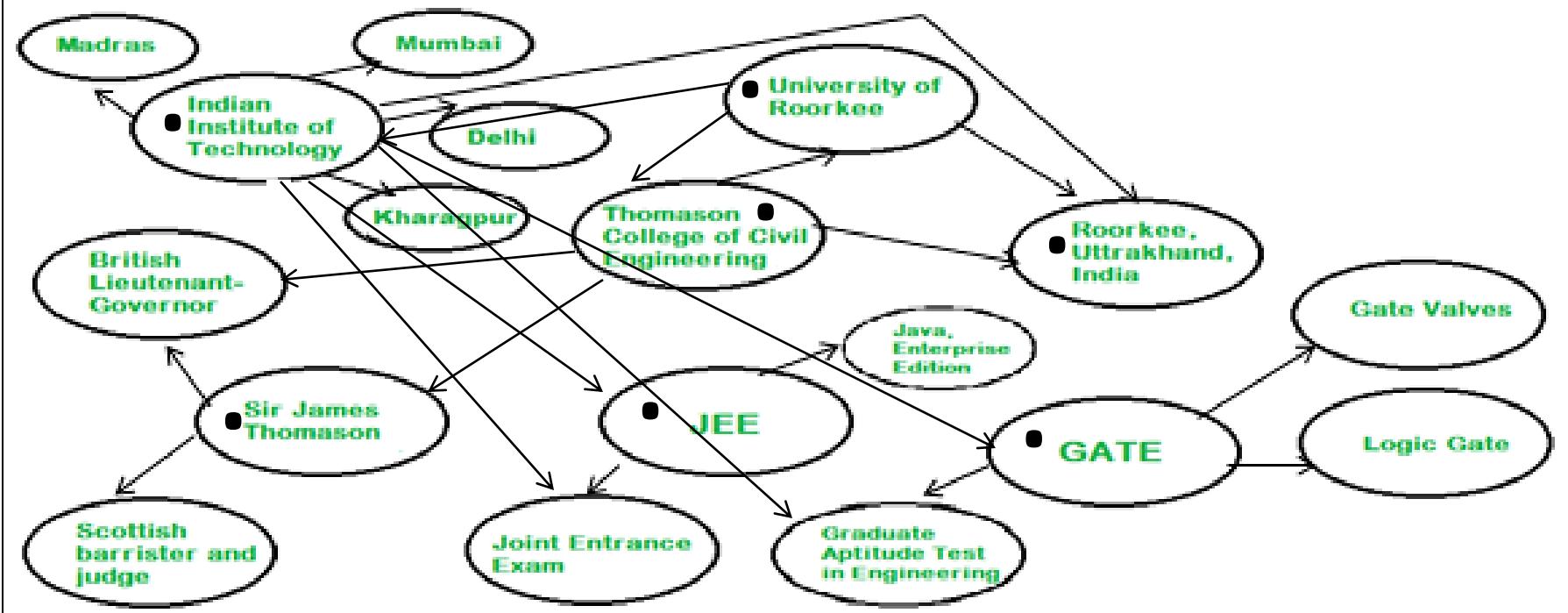
Knowledge Base

Entity Linking

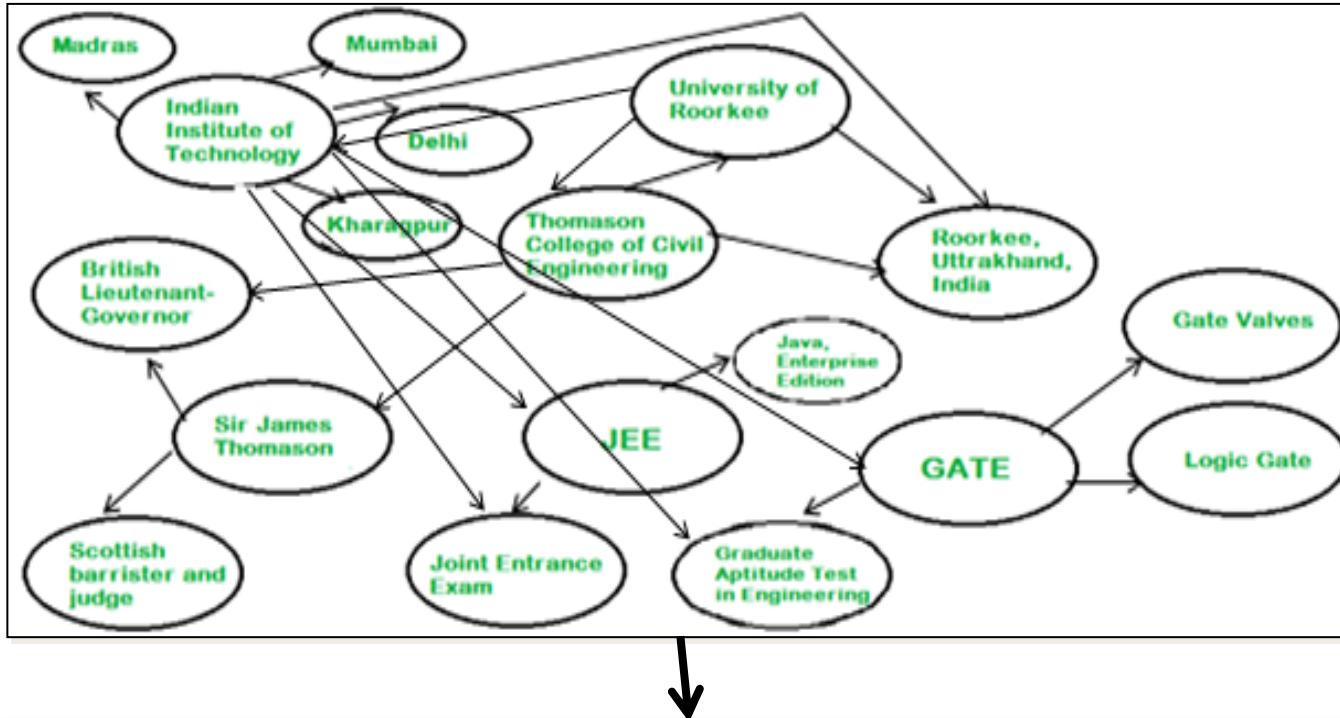
Indian Institute of Technology, formerly the University of Roorkee (1948–2001) and the Thomason College of Civil Engineering (1853–1948), is a public university located in Roorkee, Uttarakhand, India. Established in 1847 in British India by the then lieutenant governor, Sir James Thomason. The Institute admits students to B. Tech., B. Arch. and integrated M.Tech/M.Sc courses through JEE and GATE conducted at centers all over India.



Knowledge Base

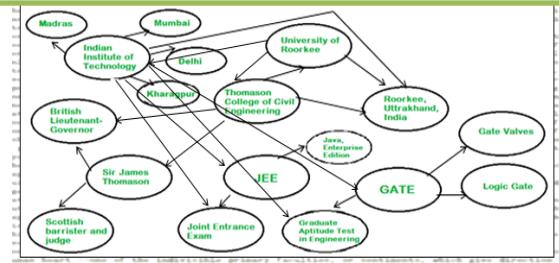


Entity Linking

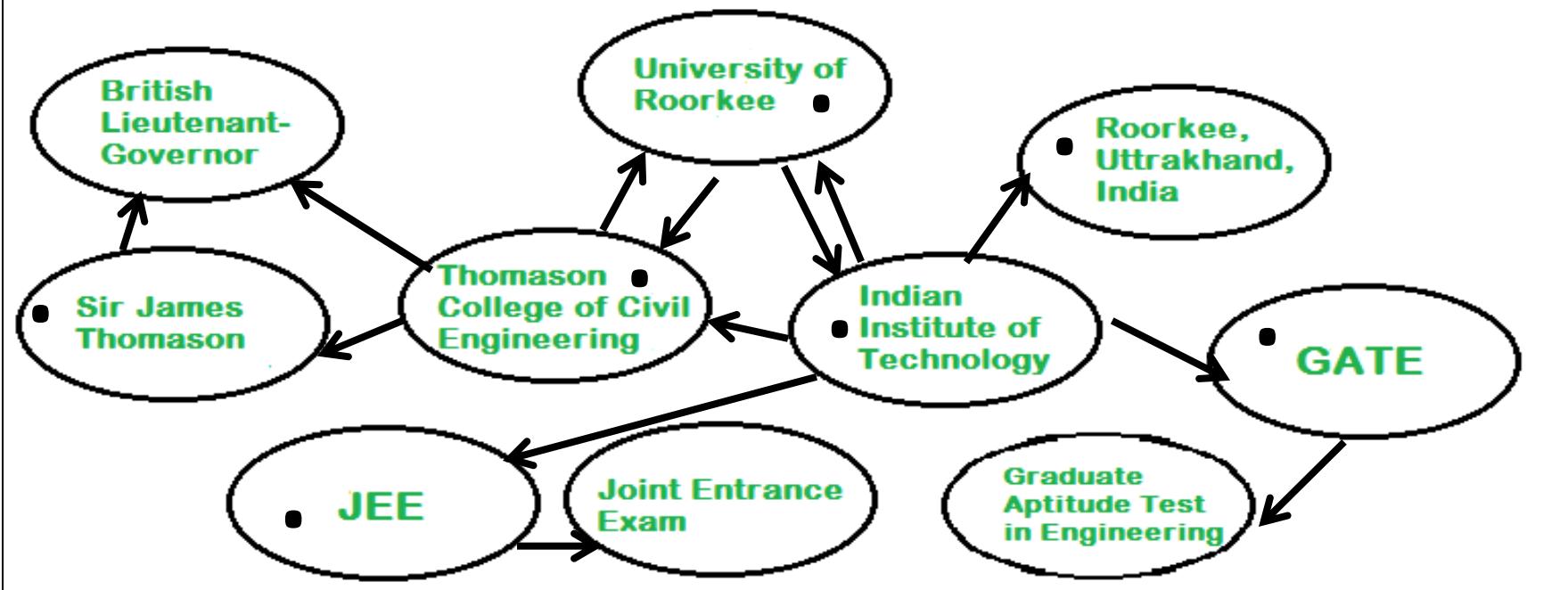


Using Random Walk with Restart

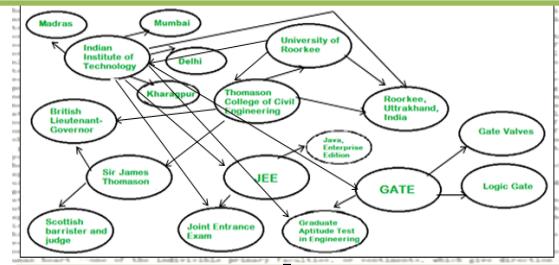
Entity Linking



Using Random Walk with Restart

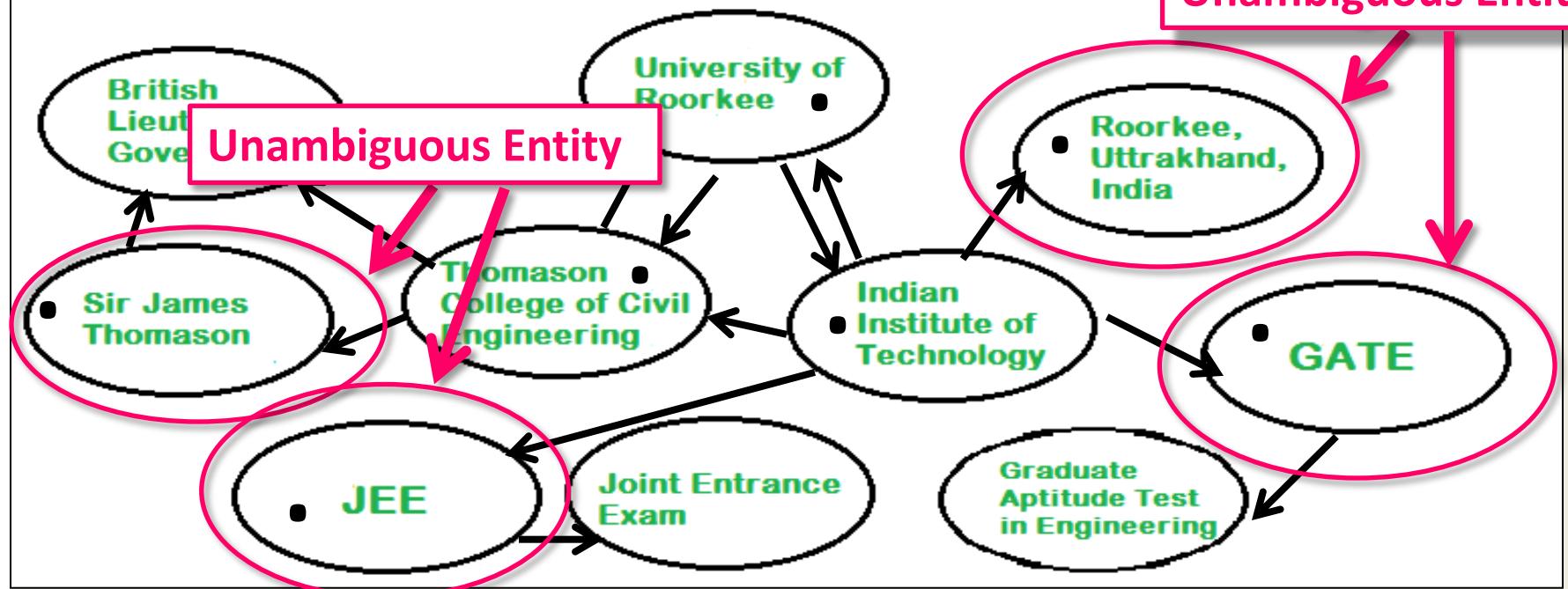


Entity Linking

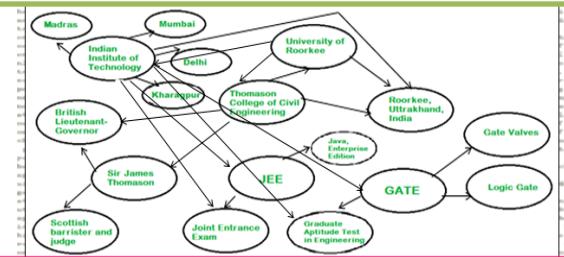


Using Random Walk with Restart

Unambiguous Entity



Entity Linking



Semantic Signature (GATE): Graduate Aptitude Test in Engineering

Semantic Signature (JEE): Joint Entrance Exam

Semantic Signature (Sir James Thomason): British Lieutenant Governor

Semantic Signature (Text): IIT, Roorkee

JEE

Joint Entrance Exam

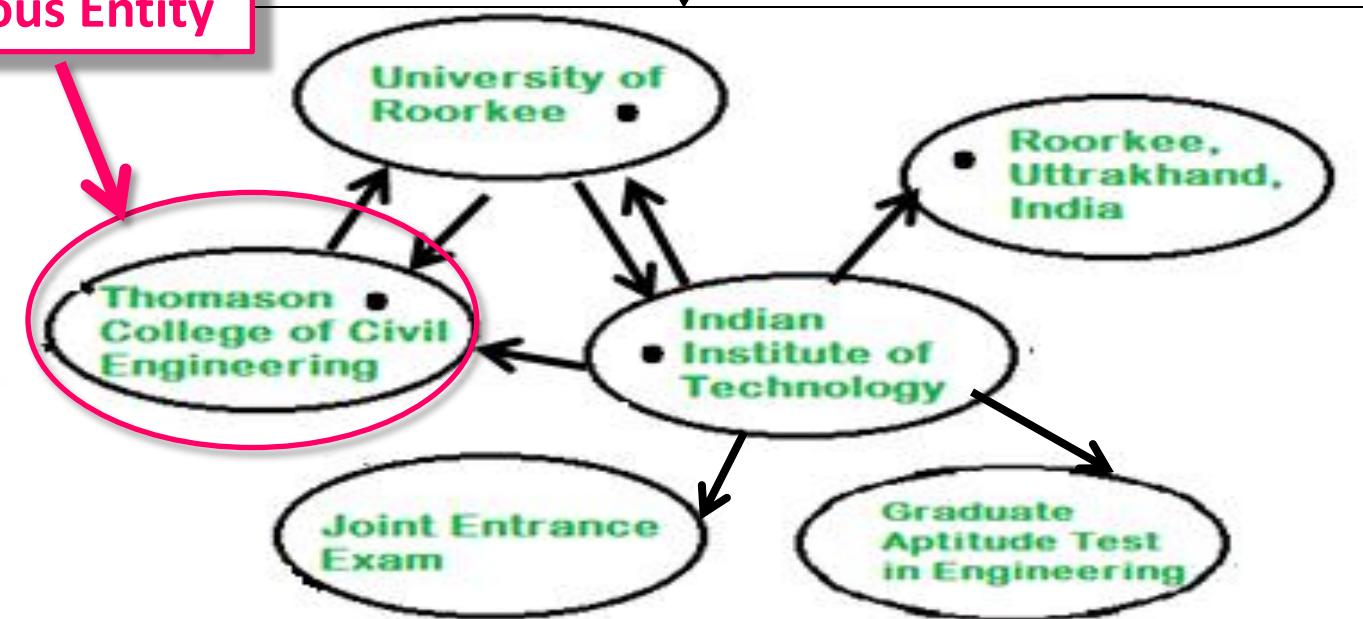
Graduate Aptitude Test in Engineering

Entity Linking

Semantic Signature (Text): IIT, Roorkee

Similarity Measure: Kullback-Leibler Divergence

Unambiguous Entity

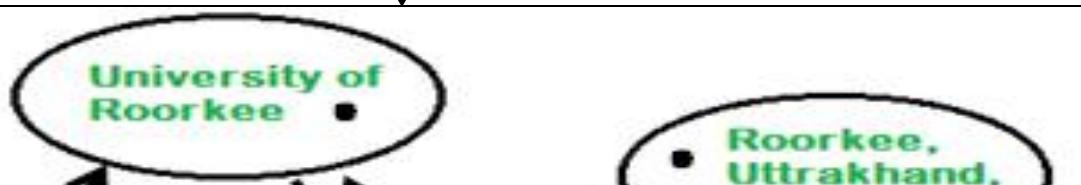


Entity Linking

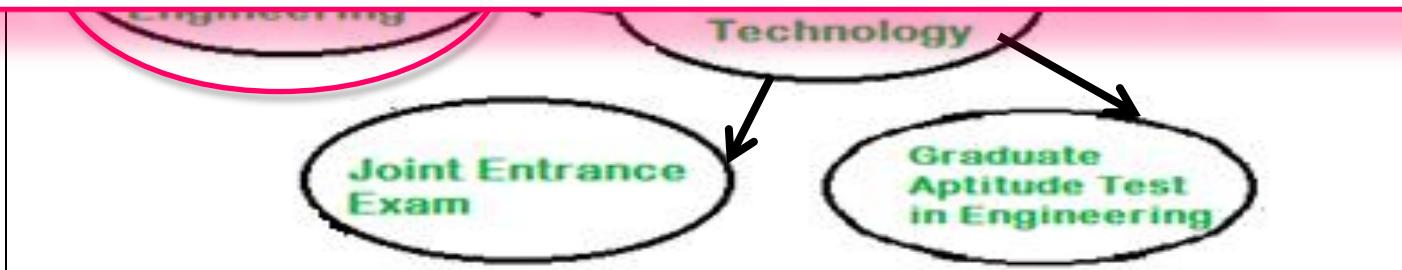
Semantic Signature (Text): IIT, Roorkee

Similarity Measure: Kullback-Leibler Divergence

Unambiguous Entity



Semantic Signature (Indian Institute of Technology): IIT Roorkee
Semantic Signature (University of Roorkee): IIT Roorkee



Entity Linking

Indian Institute of Technology, formerly the University of Roorkee (1948–2001) and the Thomason College of Civil Engineering (1853–1948), is a public university located in Roorkee, Uttarakhand, India. Established in 1847 in British India by the then lieutenant governor, Sir James Thomason. The Institute admits students to B. Tech., B. Arch. and integrated M.Tech/M.Sc courses through JEE and GATE conducted at centers all over India.



Thomason College of Civil Engineering

Indian Institute of Technology

University of Roorkee

Indian Institute of Technology
Roorkee

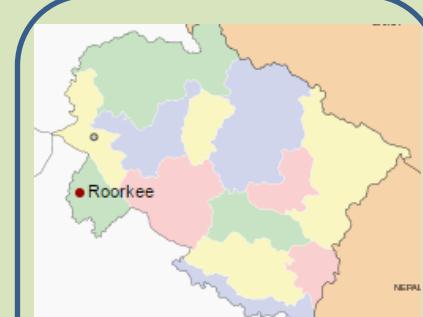
भारतीय प्रौद्योगिकी संस्थान रुडकी



Main (administrative) building



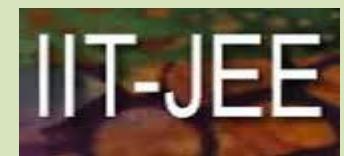
Sir James Thomason



Roorkee,
Uttarakhand, India

G.A.T.E.
Graduate Aptitude Test in Engineering

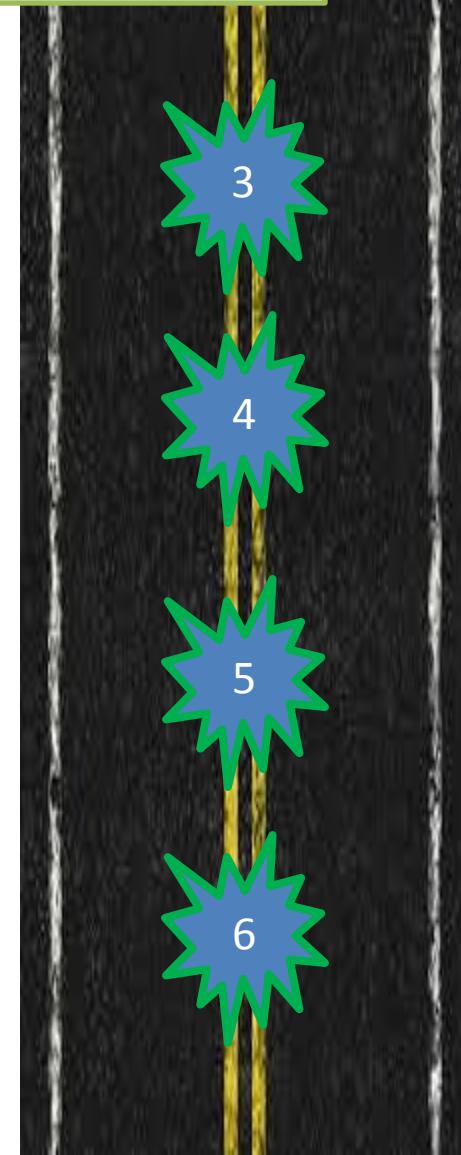
GATE



JEE

The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References

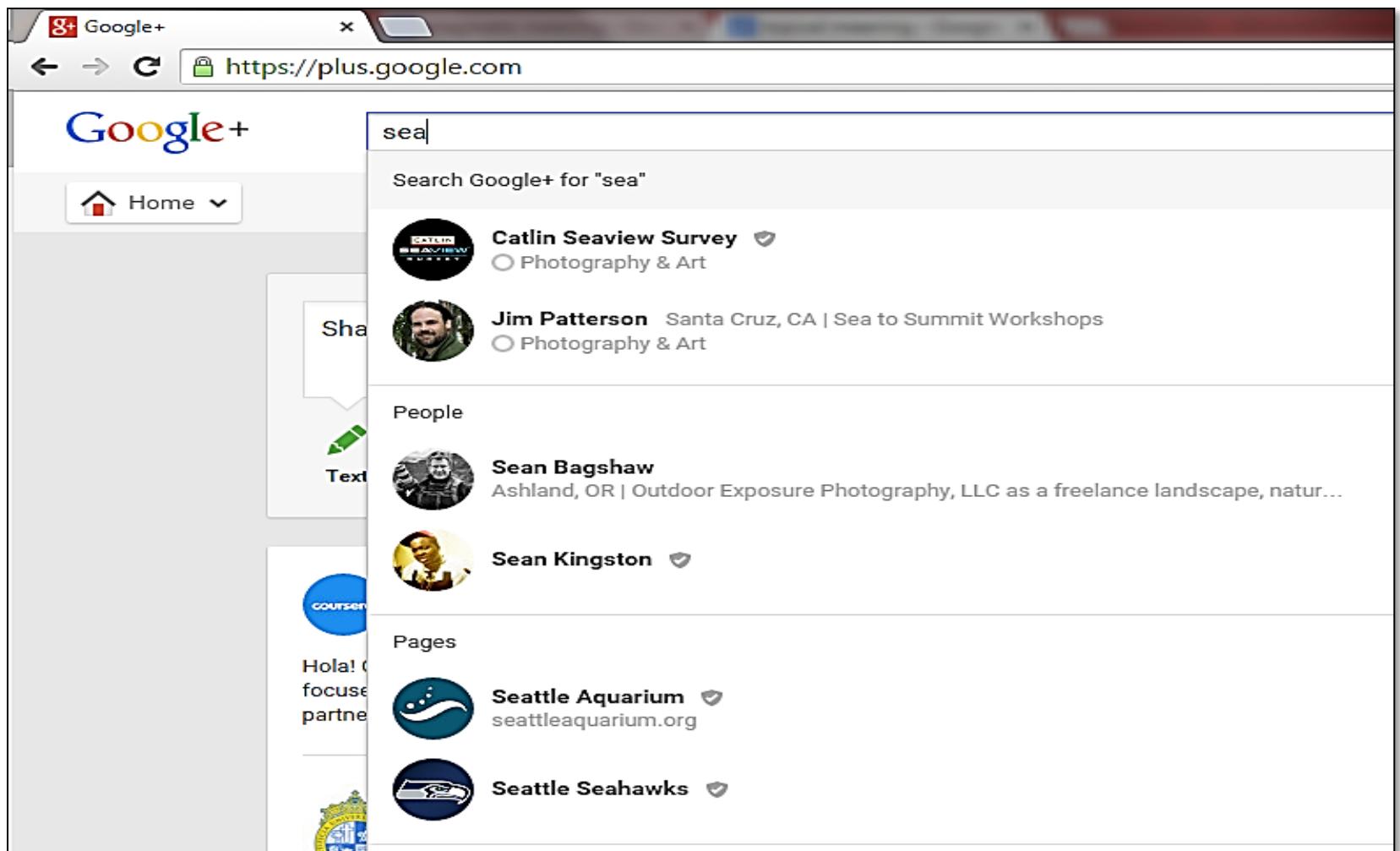


Task Extraction

- ❖ Automatic extraction of structured information from semi or un-structured raw data
- ❖ It's subtasks includes:
 - ✓ Named Entity Recognition
 - ✓ Word Sense Disambiguation
 - ✓ Coreference Resolution
 - ✓ Relationship extraction
- ❖ Some of the applications:



Task Extraction



A screenshot of a web browser showing the Google+ search interface. The URL in the address bar is <https://plus.google.com>. The search query "sea" is entered in the search bar. The results are categorized into three sections: "Search Google+ for 'sea'", "People", and "Pages".

- Search Google+ for "sea"**
 - Catlin Seaview Survey** Photography & Art
 - Jim Patterson** Santa Cruz, CA | Sea to Summit Workshops Photography & Art
- People**
 - Sean Bagshaw** Ashland, OR | Outdoor Exposure Photography, LLC as a freelance landscape, natur...
 - Sean Kingston**
- Pages**
 - Seattle Aquarium** seattleaquarium.org
 - Seattle Seahawks**

An example of **Google+** Feature based on Task Extraction

Task Extraction

What people are saying about a given entity right now?

The screenshot shows a Google+ search interface with a red border around the main query area. The URL in the address bar is https://plus.google.com. The main content area displays several search results. From left to right, there is a photo of a woman and a man in blue jerseys, the logo for Aam Aadmi Party (AAP) featuring a broom and the text 'AAP AAM AADMI PARTY', a black smartphone displaying a beach scene, and the circular emblem of the Indian Institute of Technology Roorkee. The background of the screenshot shows other Google+ search results.

An example of **Google+** Feature based on Task Extraction

Task Extraction

Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of *A Brief History of Time* which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include *A Briefer History of Time*, the essay collection *Black Holes and Baby Universe* and *The Universe in a Nutshell*.



Entity and Category Extraction Using



Task Extraction

Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of *A Brief History of Time* which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include *A Briefer History of Time*, the essay collection *Black Holes and Baby Universe* and *The Universe in a Nutshell*.



Entity and Category Extraction Using



Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of *A Brief History of Time* which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include *A Briefer History of Time*, the essay collection *Black Holes and Baby Universe* and *The Universe in a Nutshell*.

Task Extraction

Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of A Brief History of Time which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include A Briefer History of Time, the essay collection Black Holes and Baby Universe and The Universe in a Nutshell.



Entity and Category Extraction Using



Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of A Brief History of Time which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include A Briefer History of Time, the essay collection Black Holes and Baby Universe and The Universe in a Nutshell.

Entity	Category
Stephen Hawking	Researcher
Lucasian Professor of Mathematics	Researcher
University of Cambridge	Research University

Entity	Category
A Brief History of Time	Book
The Universe in a Nutshell	Book

Task Extraction

Entity	Category
Stephen Hawking	Researcher
Lucasian Professor of Mathematics	Researcher
University of Cambridge	Research University

Entity	Category
A Brief History of Time	Book
The Universe in a Nutshell	Book



Aggregation of Related Entities

Task Extraction

Entity	Category
Stephen Hawking	Researcher
Lucasian Professor of Mathematics	Researcher
University of Cambridge	Research University

Entity	Category
A Brief History of Time	Book
The Universe in a Nutshell	Book



Aggregation of Related Entities



Entity	Category
{ Stephen Hawking, Lucasian Professor of Mathematics, University of Cambridge }	Research
{ A Brief History of Time, The Universe in a Nutshell }	Book

Task Extraction

Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of A Brief History of Time which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include A Briefer History of Time, the essay collection Black Holes and Baby Universe and The Universe in a Nutshell.



Non-Entity and Category Extraction

Task Extraction

Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of A Brief History of Time which was an international bestseller. Now the Dennis Stanton Avery and Sally Tsui Wong-Avery Director of Research at the Department of Applied Mathematics and Theoretical Physics and Founder of the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include A Briefer History of Time, the essay collection Black Holes and Baby Universe and The Universe in a Nutshell.



Non-Entity and Category Extraction



Non-Entity	Category	Non-Entity	Category
Theoretical Physics	Research	Author	Book
Applied Mathematics	Research	International Bestseller	Book
Theoretical Cosmology	Research	General Reader	Book

Task Extraction

Entity	Category
{ Stephen Hawking, Lucasian Professor of Mathematics, University of Cambridge }	Research
{ A Brief History of Time, The Universe in a Nutshell }	Book

Non-Entity	Category
{Theoretical Physics, Applied Mathematics, Theoretical Cosmology}	Research
{ Author, International Bestseller, General Reader}	Book



Associate Non-Entity and Entity with respect to Category

Task Extraction

Entity	Category
{ Stephen Hawking, Lucasian Professor of Mathematics, University of Cambridge }	Research
{ A Brief History of Time, The Universe in a Nutshell }	Book

Non-Entity	Category
{Theoretical Physics, Applied Mathematics, Theoretical Cosmology}	Research
{ Author, International Bestseller, General Reader}	Book



Associate Non-Entity and Entity with respect to Category



Entity	Non-Entity	Category
Stephen Hawking, Lucasian Professor of Mathematics, University of Cambridge	Theoretical Physics, Applied Mathematics, Theoretical Cosmology	Research
A Brief History of Time, The Universe in a Nutshell	Author, International Bestseller, General Reader	Book

Task Extraction

Entity	Non-Entity	Category
Stephen Hawking, Lucasian Professor of Mathematics, University of Cambridge	Theoretical Physics, Applied Mathematics, Theoretical Cosmology	Research
A Brief History of Time, The Universe in a Nutshell	Author, International Bestseller, General Reader	Book



Ranking each list within a Category Using
Term Frequency- Inverse Document Frequency

Task Extraction

Entity	Non-Entity	Category
Stephen Hawking, Lucasian Professor of Mathematics, University of Cambridge	Theoretical Physics, Applied Mathematics, Theoretical Cosmology	Research
A Brief History of Time, The Universe in a Nutshell	Author, International Bestseller, General Reader	Book



Ranking each list within a Category Using
Term Frequency- Inverse Document Frequency



Entity	Non-Entity	Category
University of Cambridge, Stephen Hawking, Lucasian Professor of Mathematics	Applied Mathematics, Theoretical Physics, Theoretical Cosmology	Research
A Brief History of Time, The Universe in a Nutshell	International Bestseller, Author, General Reader	Book

Task Extraction

Entity	Non-Entity	Category
University of Cambridge, Stephen Hawking, Lucasian Professor of Mathematics	Applied Mathematics, Theoretical Physics, Theoretical Cosmology	Research
A Brief History of Time, The Universe in a Nutshell	International Bestseller, Author, General Reader	Book



Pruning items within each list having frequency below threshold (pre defined in problem)

Task Extraction

Entity	Non-Entity	Category
University of Cambridge, Stephen Hawking, Lucasian Professor of Mathematics	Applied Mathematics, Theoretical Physics, Theoretical Cosmology	Research
A Brief History of Time, The Universe in a Nutshell	International Bestseller, Author, General Reader	Book



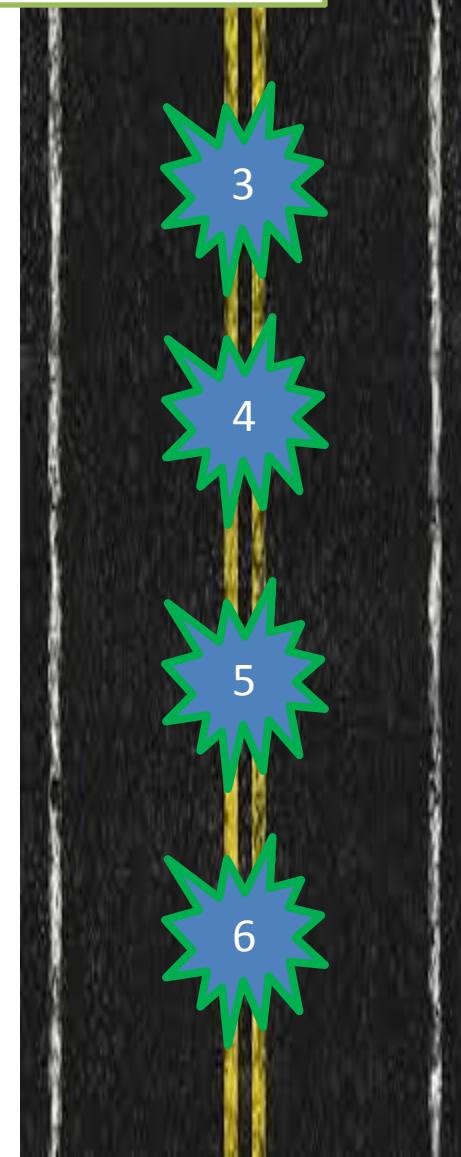
Pruning items within each list having frequency below threshold (pre defined in problem)



Entity	Non-Entity	Category
University of Cambridge, Stephen Hawking	Applied Mathematics, Theoretical Physics	Research
A Brief History of Time, The Universe in a Nutshell	International Bestseller, Author	Book

The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



Gain

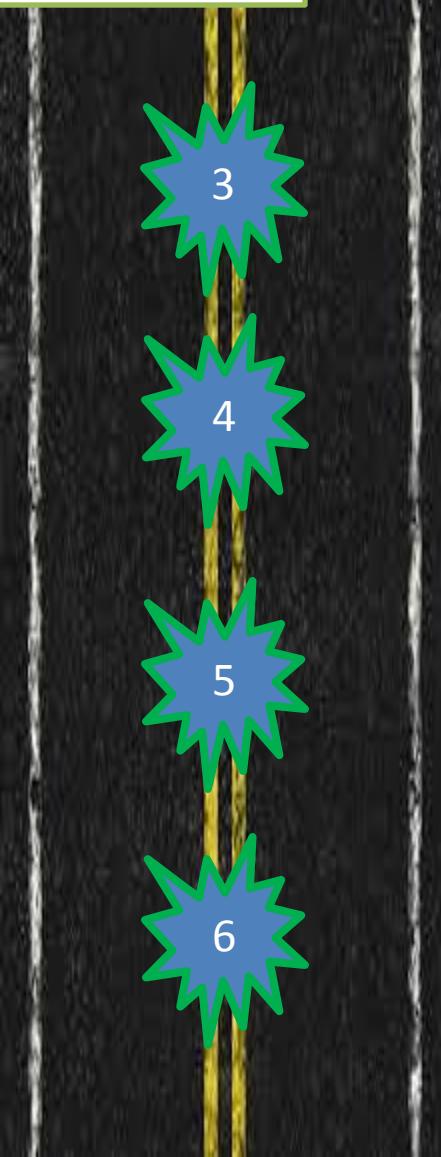


- Now, we know how to reduce the burden of Mr. X, by leveraging computational machine.
- We also got to know what are the things we need to do, to make machine able to (process + understand + synthesize) Natural/Human Language
- We know how to gain knowledge from semi/unstructured *informational sources.



The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
- 8. Future Research Direction**
9. References



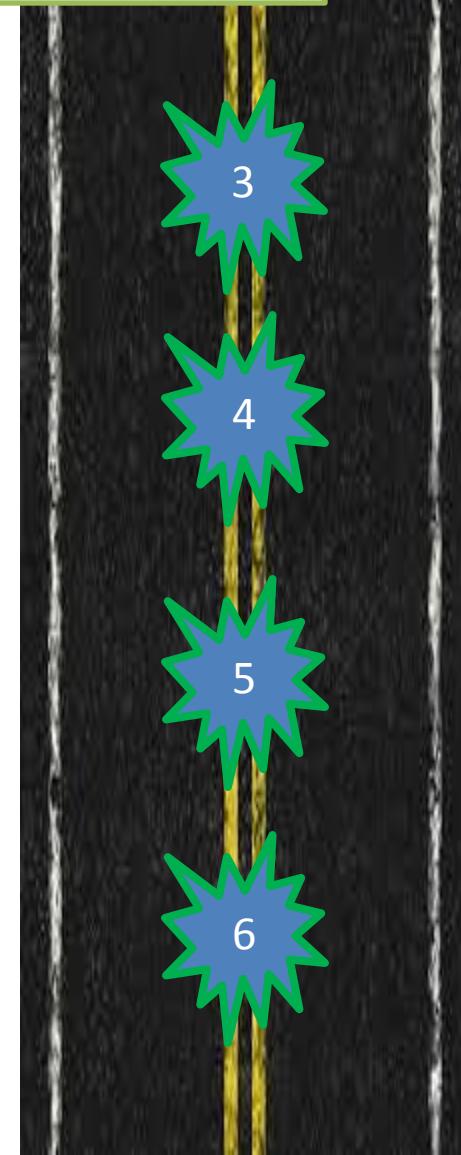
Future Work



- Information extraction using Image, Videos, Sound tracks.
- Automatic creation of idiosyncratic knowledge bases.
- Disambiguation of news articles.

The Roadmap

1. Inspiration
2. How to make machine smart?
3. Entity Recognition
4. Entity Disambiguation
5. Entity Linking
6. Entity Related Task Extraction
7. Gain
8. Future Research Direction
9. References



References

- Roman Prokofyev, Gianluca Demartini, and Philippe Cudré-Mauroux , “Effective Named Entity Recognition for Idiosyncratic Web Collections,”
- Andrea Moro, Alessandro Raganato, Roberto Navigli, “Entity Linking meets Word Sense Disambiguation: a Unified Approach,”
- “Dexter”, An Open Source Framework for Entity Linking, [Online Demo]. Available: <http://dexter.isti.cnr.it/demo/>. [Accessed April, 2015].
- “Babelfy”, Unified, Multilingual, graph based approach to Entity Linking and Word Sense Disambiguation, [Online Demo]. Available: <http://babelfy.org/index.jsp>. [Accessed April, 2015].
- Luciano Del Corro, Rainer Gemulla and Gerhard Weikum, “Werdy: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning,”
- Manisha Verma and Emine Yilmaz, “Entity oriented Task Extraction from Query Logs,”
- "Google," Google Inc., [Online]. Available: <https://www.google.co.in/>. [Accessed April, 2015].

for the most wild, yet most humbly narrative which I am about to pen. I neither expect nor call it belief. Had indeed could I be to expect it, in a case where my very senses reject their own evidence. Yet, and as I not --and very surely do I not dream. But to-morrow I die, and to-day I would unburden my soul. My immediate purpose is to place before the world, plainly, succinctly, and without comment, a series of mere household events. In their consequences, these events have terrified --have tortured --have destroyed me. Yet I will not attempt to expound them. To me, no scene is more terrible than banqueting. But phantoms to the common-place than my own, which will personify an ordinary succession of every triviality and banality of my die

A cartoon illustration of a blue bird with large white eyes and a yellow beak. It is holding a light brown rectangular sign with the words "Thank you!" written on it in a stylized font. The bird's hands are yellow and it has a small tuft of blue feathers on its head. The entire image is framed by a red border.

Questions?