

Semantic Web: Mining Knowledge from the Web

Raghava Mutharaju

Knowledgeable Computing and Reasoning Lab

IIT-Delhi



About Myself

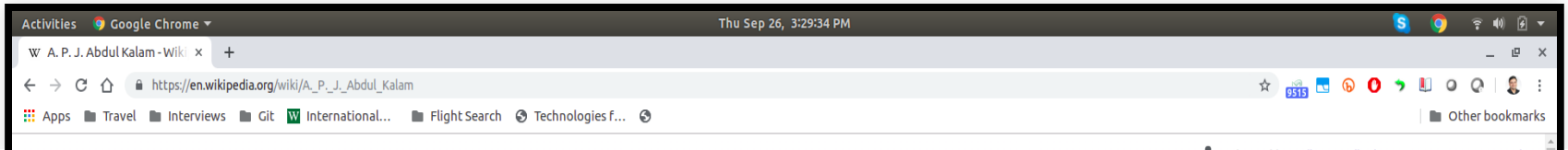
Work Experience and Education

- Assistant Professor (CSE), IIIT-D
 - Research Scientist, GE Research Center, New York
 - Internships at IBM Research, Bell Labs, Xerox Research, Stardog
 - Software Engineer, CA Technologies, Hyderabad
-
- PhD from Wright State University
 - M.Tech from MNNIT, Allahabad
 - B.Tech from JNTU, Hyderabad

Research Interests

- Semantic Web
 - Ontology Modelling
 - Ontology Reasoning
 - Knowledge Graphs
 - SPARQL Query Processing
- Big Data
- IoT

Introduction



1. Who was the president of India who was also a scientist?
2. Who was the scientist who worked at DRDO and was a Bharat Ratna?
3. Who was the president of India with an aerospace engineering background?
4. Name the person that worked at DRDO and ISRO and was called the “People’s President”?
5. Who was the Bharat Ratna holder from rameswaram?

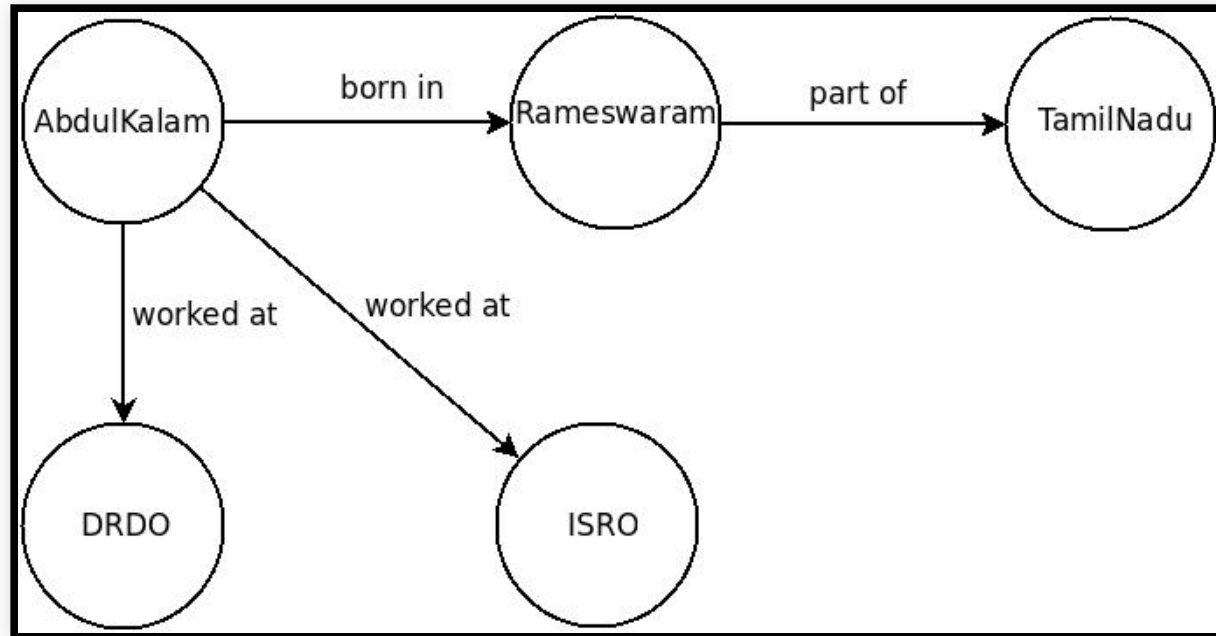
Humans

- can read the text
- can understand the text
- can make inferences based on the text
- can answer the questions

Machines

- can read the text
- cannot understand the text
- cannot make any inferences
- may answer some of the questions

Knowledge Graph

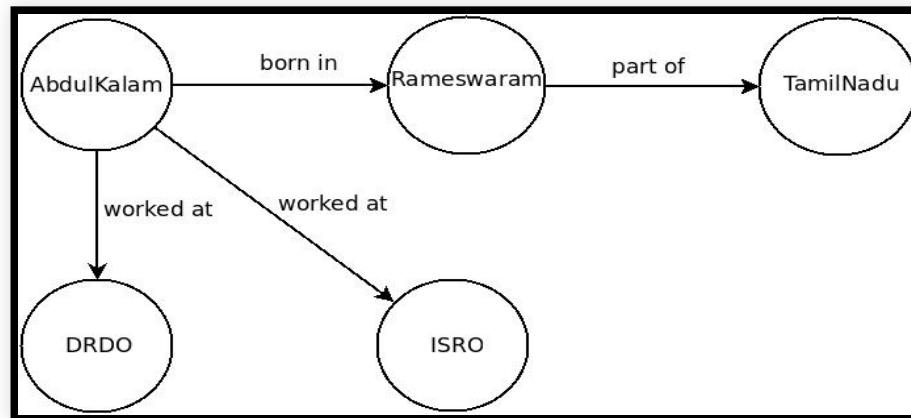


Knowledge Graph involving Abdul Kalam

Knowledge Graph

- Capture the knowledge in a structured form
- Machine processable
- Inferences can be drawn
- KG can be linked to other related KGs
 - Presidents of India
 - Books written by presidents
 - Details of DRDO and ISRO
- Machines have better chance to answer questions using KGs

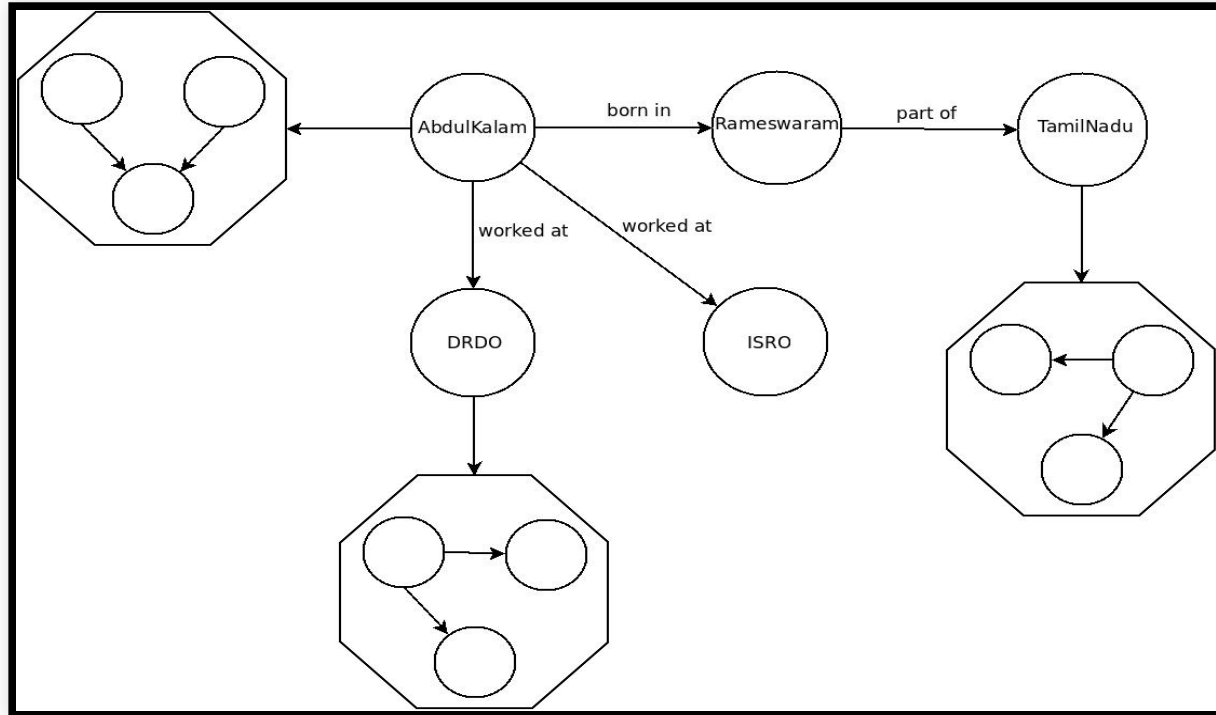
Knowledge Graph



Knowledge Graph involving Abdul Kalam

Domain(born in, Person) Range(born in, Place) Range(worked at, Organization)

Knowledge Graph



Expanded Knowledge Graph involving Abdul Kalam

Activities Google Chrome Sat Sep 28, 12:03:17 AM


abdul kalam - Google Search

https://www.google.com/search?safe=active&sxsrf=ACYBGNRQ3MIQXKTFnETXED_29VzrW_NH0g%3A1569609093206&source=hp&ei=hVWOXcPJCsH9QOMjYa4Aw&q=abdul+kalam&oq=abdul+kalam&gs_l=...

Apps Travel Interviews Git International... Flight Search Technologies f... Other bookmarks

Google abdul kalam

Images for abdul kalam



More images for abdul kalam Report images

A. P. J. Abdul Kalam - Wikipedia

https://en.wikipedia.org/wiki/A._P._J._Abdul_Kalam

Avul Pakir Jainulabdeen Abdul Kalam was an aerospace scientist who served as the 11th President of India from 2002 to 2007. He was born and raised in ...

Vice President: [Krishan Kant; Bhairon Singh S...](#) Prime Minister: [Atal Bihari Vajpayee; Manmoh...](#)

Resting place: [Dr. A. P. J. Abdul Kalam National ...](#) Profession: [Aerospace scientist; Author](#)

[Dr. APJ Abdul Kalam Award · Wings of Fire · I Am Kalam · Defence Research](#)

You visited this page on 19/9/19.

[Dr.APJ.Abdulkalam Official Website | Dr.Kalam Speeches | Dr ...](#)

www.abdulkalam.com

Kerala Opens Museum Dedicated to Dr.A.P.J.Abdul Kalam - First of its kind in South India.
Kerala gave a fitting tribute to India's former President Dr. A P J Abdul ...

A. P. J. Abdul Kalam

Former President of India

Avul Pakir Jainulabdeen Abdul Kalam was an aerospace scientist who served as the 11th President of India from 2002 to 2007. He was born and raised in Rameswaram, Tamil Nadu and studied physics and aerospace engineering. [Wikipedia](#)

Born: 15 October 1931, [Rameswaram](#)

Died: 27 July 2015, [Shillong](#)


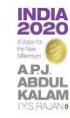
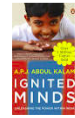
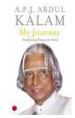

Full name: [Avul Pakir Jainulabdeen Abdul Kalam](#)

Awards: [Bharat Ratna, Padma Bhushan, Padma Vibhushan, MORE](#)

Education: [Madras Institute Of Technology, Anna University \(1955–1960\), MORE](#)






Books

View 45+ more

				
Agni Ki Udaan	India 2020	Ignited Minds	My Journey: Transformations	Turning Points: A Journey Through
1999	1998	2002	2013	2012

People also search for

View 15+ more

				
Bhagat Singh	C. V. Raman	Vikram Sarabhai	Pratibha Patil	Homi J. Bhabha

Google's knowledge card for Abdul Kalam

Knowledge Graph

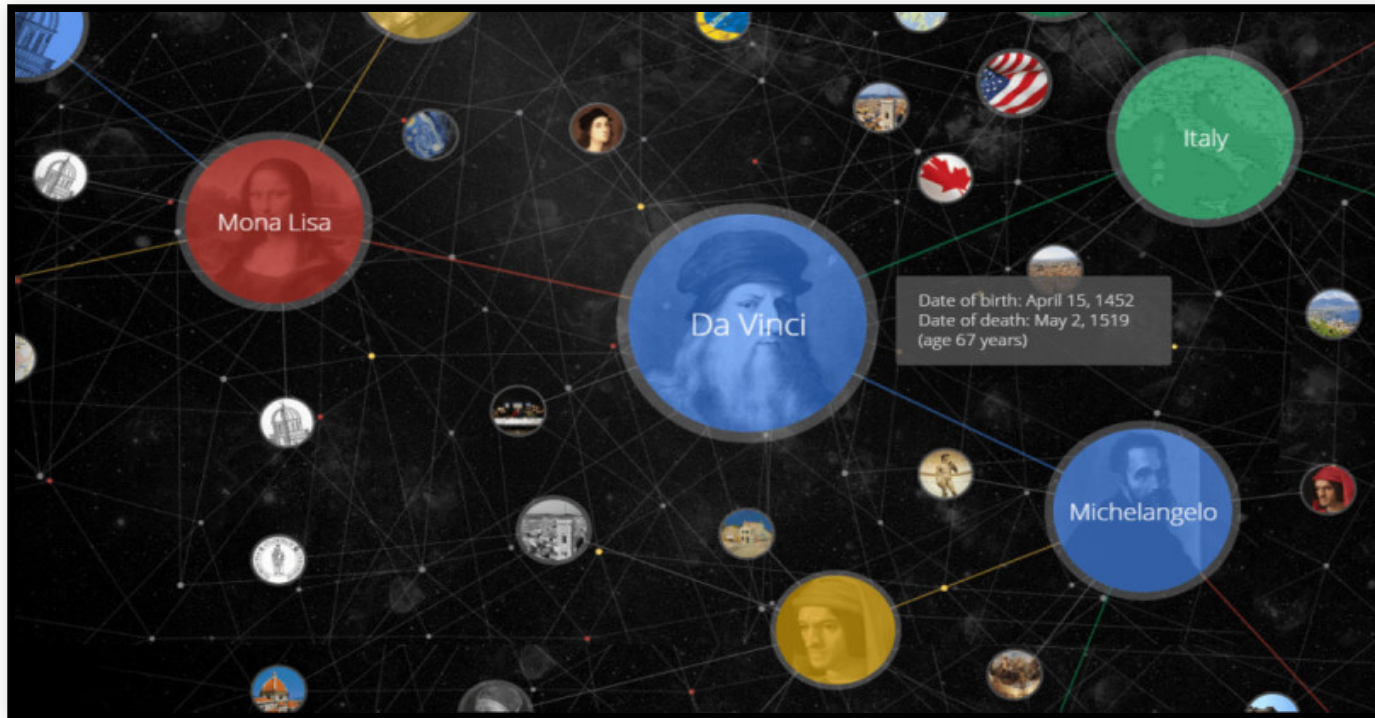


Image source: <https://goo.gl/S2F3mH>

Semantic Web

- Purpose is to provide structure to the Web and to the data in general
- Move from web of documents to web of data
- Proposed by Tim-Berners Lee in 1999
- Semantic Web technologies and W3C standards
 - *RDF (Resource Description Framework)*
 - OWL (Web Ontology Language)
 - SPARQL (query language)
 - SHACL (Shapes Constraint Language)

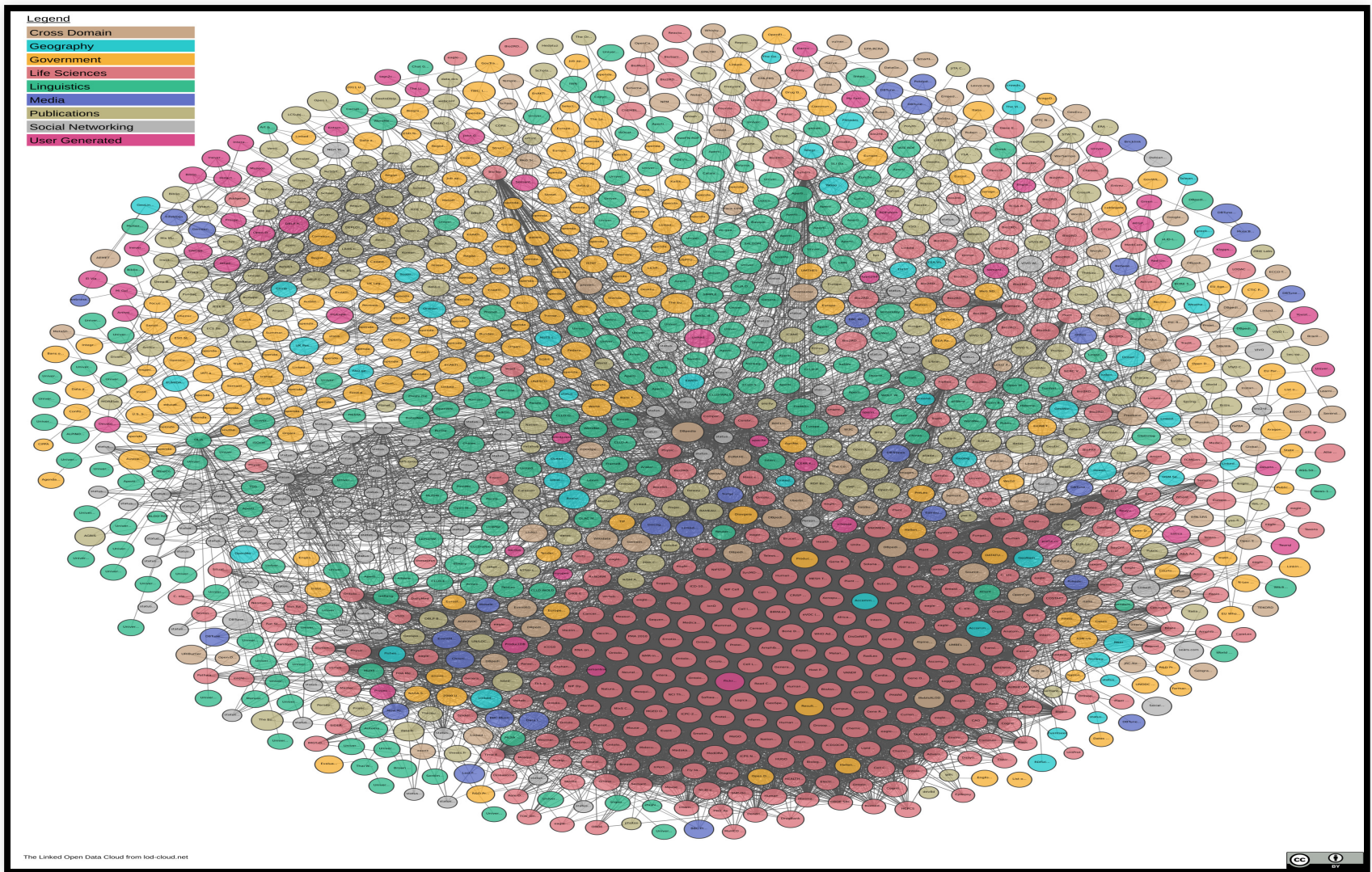


Image source: <https://lod-cloud.net/>

RDF

- Resource Description Framework (RDF) is a data model that is used to describe resources
 - Physical things
 - Abstract concepts
 - Numbers and strings
- We use the term resource and entity synonymously here
- It is a universal, machine readable data exchange format
- Resources are described using triples (subject, predicate, object)
- A triple captures the relationship (predicate) between a subject and an object

- <Delhi> <capitalOf> <India>
- RDF triple is called a RDF Statement
- Triple can be represented as a directed labelled graph

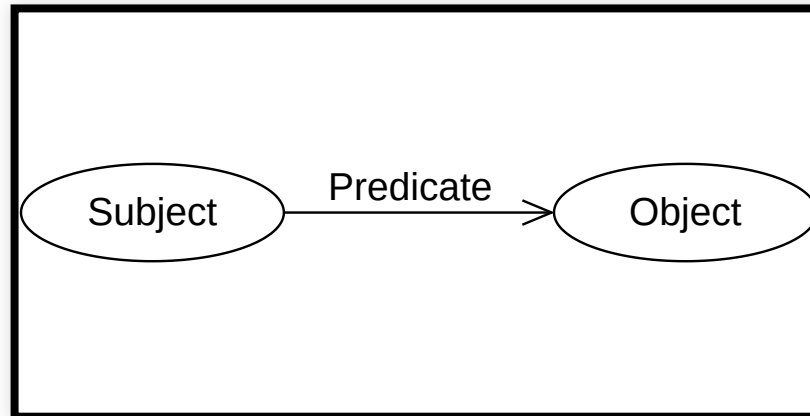


Image source: <https://www.w3.org/TR/rdf11-concepts/rdf-graph.svg>

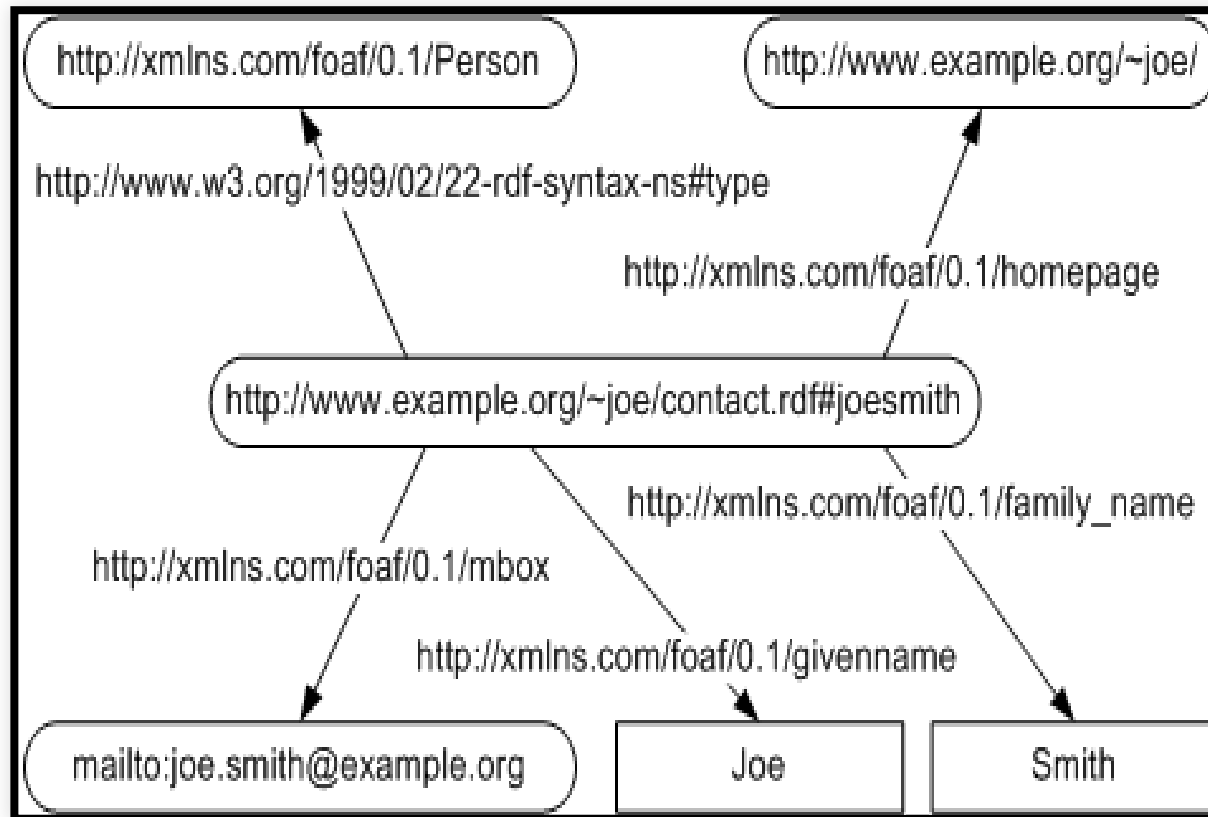


Image source: <http://www.obitko.com/tutorials/ontologies-semantic-web/rdf-graph-and-syntax.html>

Mining Knowledge

Activities Google Chrome Thu Sep 26, 3:29:34 PM


W A. P. J. Abdul Kalam - Wiki x +


https://en.wikipedia.org/wiki/A._P._J._Abdul_Kalam

Apps Travel Interviews Git International... Flight Search Technologies f... Other bookmarks

Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

 Wiki Loves Monuments: Photograph a monument, help Wikipedia and win!
[Learn more](#)



A. P. J. Abdul Kalam

From Wikipedia, the free encyclopedia

*In this Indian name, the name Avul Pakir Jainulabdeen is a **patronymic**, not a **family name**, and the person should be referred to by the **given name**, Abdul Kalam.*

Avul Pakir Jainulabdeen Abdul Kalam (/æbdʊl kəlɑːm/ (listen); 15 October 1931 – 27 July 2015) was an aerospace scientist who served as the **11th President of India** from 2002 to 2007. He was born and raised in **Rameswaram, Tamil Nadu** and studied physics and aerospace engineering. He spent the next four decades as a scientist and science administrator, mainly at the **Defence Research and Development Organisation** (DRDO) and **Indian Space Research Organisation** (ISRO) and was intimately involved in India's civilian space programme and military **missile development efforts**.^[1] He thus came to be known as the *Missile Man of India* for his work on the development of **ballistic missile** and **launch vehicle** technology.^{[2][3][4]} He also played a pivotal organisational, technical, and political role in India's *Pokhran-II* nuclear tests in 1998, the first since the **original nuclear test by India** in 1974.^[5]


Kalam was **elected** as the 11th President of India in 2002 with the support of both the ruling **Bharatiya Janata Party** and the then-opposition **Indian National Congress**. Widely referred to as the "People's President",^[6] he returned to his civilian life of education, writing and public service after a single term. He was a recipient of several prestigious awards, including the **Bharat Ratna**, India's highest civilian honour.

While delivering a lecture at the **Indian Institute of Management Shillong**, Kalam collapsed and died from an apparent cardiac arrest on 27 July 2015, aged 83.^[7] Thousands including national-level dignitaries attended the funeral ceremony held in his hometown of Rameshwaram, where he was buried with **full state honours**.^[8]

Contents [hide]

- Early life and education
- Career as a scientist
- Presidency
- Post-presidency

A. P. J. Abdul Kalam



11th President of India

In office

25 July 2002 – 25 July 2007

Prime Minister [Atal Bihari Vajpayee](#)
[Manmohan Singh](#)

Vice President [Krishan Kant](#)
[Bhairon Singh Shekhawat](#)

Preceded by [K. R. Narayanan](#)

Knowledge Graph Construction

- Manual
 - Good quality KG (entities and relations)
 - Not scalable
- Automatic
 - Quality unpredictable, but generally not good
 - Scalable: large amount of text, any domain

Automatic Knowledge Graph Construction

- Steps
 - Parts of speech tagging
 - Identify nouns, verbs, adjectives, etc.
 - Dependency parsing
 - Recognize the syntactic structure of the sentence along with the dependencies among the words
 - Named Entity Recognition
 - Identify person, location, organization, etc.

Automatic Knowledge Graph Construction

- Steps
 - Coreference resolution
 - All words that refer to the same entity (I, we, he, she, it)
 - Entity resolution and linking
 - Uniquely identify entities that refer to the same entity
 - Obama, Barack Obama, President, Paris, Amazon
 - Information extraction
 - Define the domain (high level concepts and relations)
 - Learn extractors/templates
 - Rank/score the candidate facts

Automatic Knowledge Graph Construction

Existing OpenIE Systems

- OpenIE 5.0
- FRED
- ClausIE
- MinIE

- Demo

List of papers/code/data

- <https://github.com/gkiril/oie-resources>

Challenges

- Large scale benchmark for all the OpenIE systems. Results should be reproducible.
- OpenIE systems for non-English languages
- Handle all types of sentences from different domains
 - Long/short sentences, with clauses, numerical values, etc.

Semantic Web: Mining Knowledge from the Web

Conclusion

- In order for the machines to understand and make sense of the data, it has to be in a structured form
- Knowledge can be captured and (machine) processed if it is in the form of Knowledge Graphs
- Automatically building Knowledge Graphs is hard
 - Quality of the triples is not good
 - Some of the open challenges were discussed

