

Artificial Intelligence: Representation of Knowledge & Beyond Part - 2



Niladri Chatterjee, Ph.D.
Chair Professor of Artificial Intelligence
Indian Institute of Technology Delhi
Email: niladri@maths.iitd.ac.in

Semantic Annotation

Semantic annotation is the process of attaching additional information to various concepts (e.g., people, things, places, organizations, etc.) in a given text or any other content.

<https://www.ontotext.com/knowledgehub/fundamentals/semantic-annotation/>

Semantic annotation is the process of attaching additional information to various concepts (e.g., people, things, places, organizations, etc.) in a given text or any other content.

When a document is semantically annotated, it becomes a source of information that is easy to interpret, combine and reuse by our computers.

<https://www.ontotext.com/knowledgehub/fundamentals/semantic-annotation/>

Semantic annotation is the process of attaching additional information to various concepts (e.g., people, things, places, organizations, etc.) in a given text or any other content.

When a document is semantically annotated, it becomes a source of information that is easy to interpret, combine and reuse by our computers.

We can use annotated text in conjunction with expert/domain knowledge to build the knowledge base needed for developing an Information system.

Semantic annotation is the process of attaching additional information to various concepts (e.g., people, things, places, organizations, etc.) in a given text or any other content.

When a document is semantically annotated, it becomes a source of information that is easy to interpret, combine and reuse by our computers.

We can use annotated text in conjunction with expert/domain knowledge to build the knowledge base needed for developing an Information system.

For illustration consider the following:

NC and PS teach AI, also known as Artificial Intelligence, in IIT Delhi for B.Tech and M.Tech programmes in Comp. Sc.

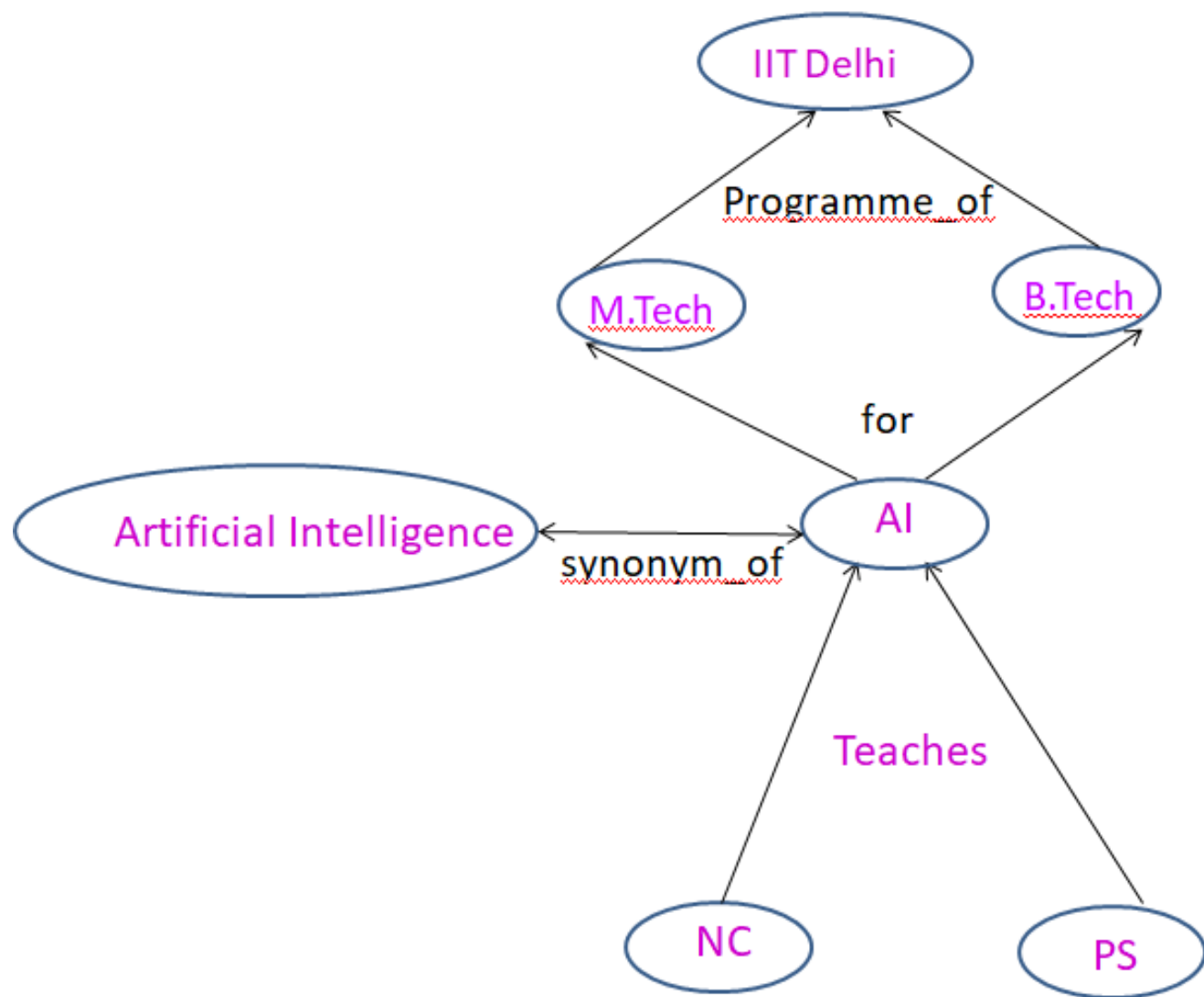
This sentence may be annotated as:

<NC: person name> and <PS : person name> teach <AI : subject name > , also known as
<Artificial_Intelligence: subject name>
in <IIT_Delhi: institute> for <B.Tech: program name > and <M.Tech: program name >
programmes in <Comp. Sc.: domain name>

This sentence may be annotated as:

<NC: **person name**> and <PS : **person name**> teach <AI : **subject name** > , also known as
<Artificial_Intelligence: **subject name**>
in <IIT_Delhi: **institute**> for <B.Tech: **program name** > and <M.Tech: **program name** >
programmes in <Comp. Sc.: **domain name**>

We may design the following graphical representation for the above sentence



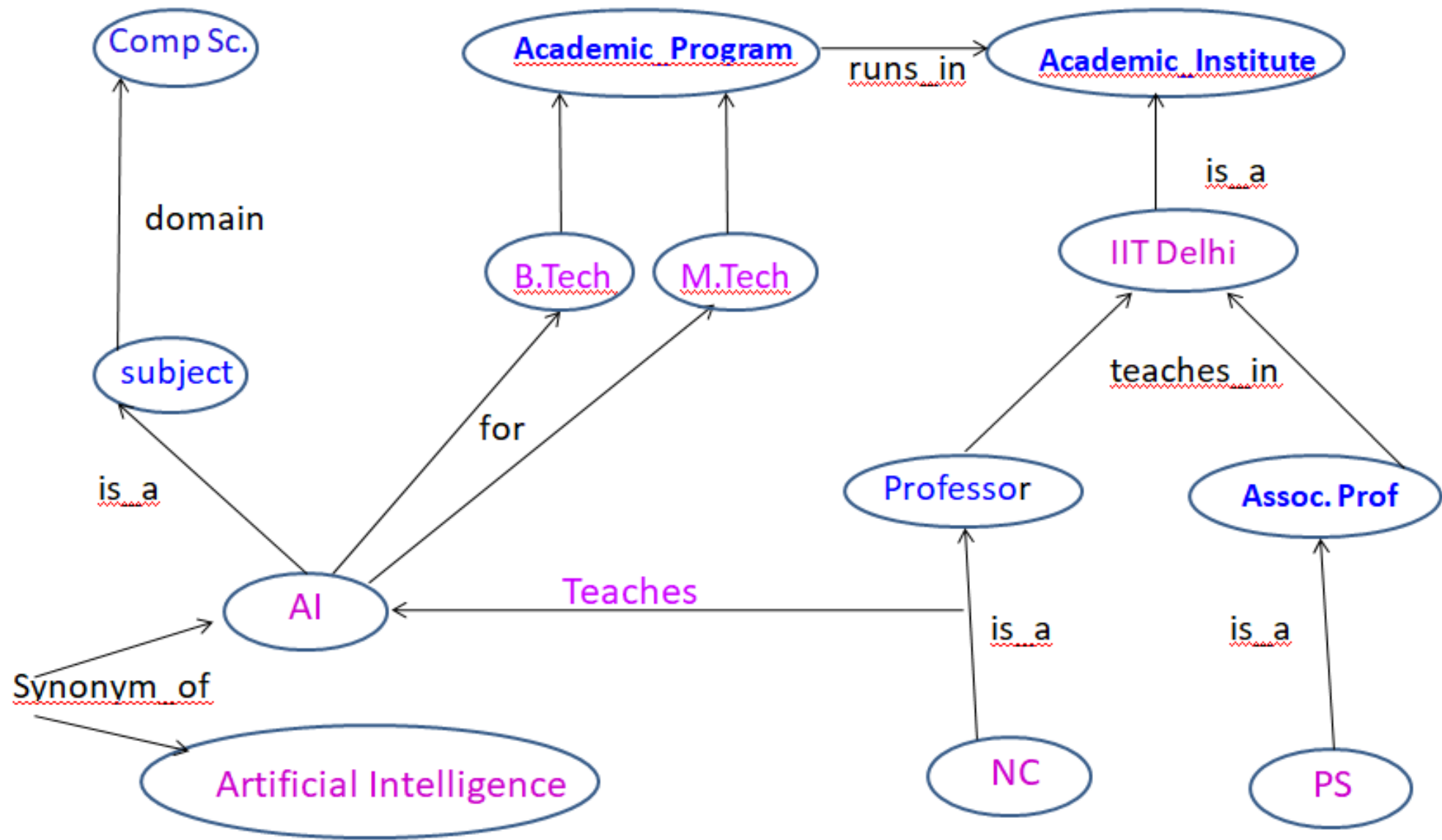
But this graph does not help in finding answers of some domain-based questions :

- * Who is NC ?
- * What is IIT Delhi ?
- * What is B.Tech ?
- * Who studies AI ?
- * What is Artificial Intelligence ?

But this graph does not help in finding answers of some domain-based questions :

- * Who is NC ?
- * What is IIT Delhi ?
- * What is B.Tech ?
- * Who studies AI ?
- * What is Artificial Intelligence ?

We may use further domain knowledge to fill in these gaps so that certain more general questions can be answered.



Still it is not interoperable with inter-domain queries:

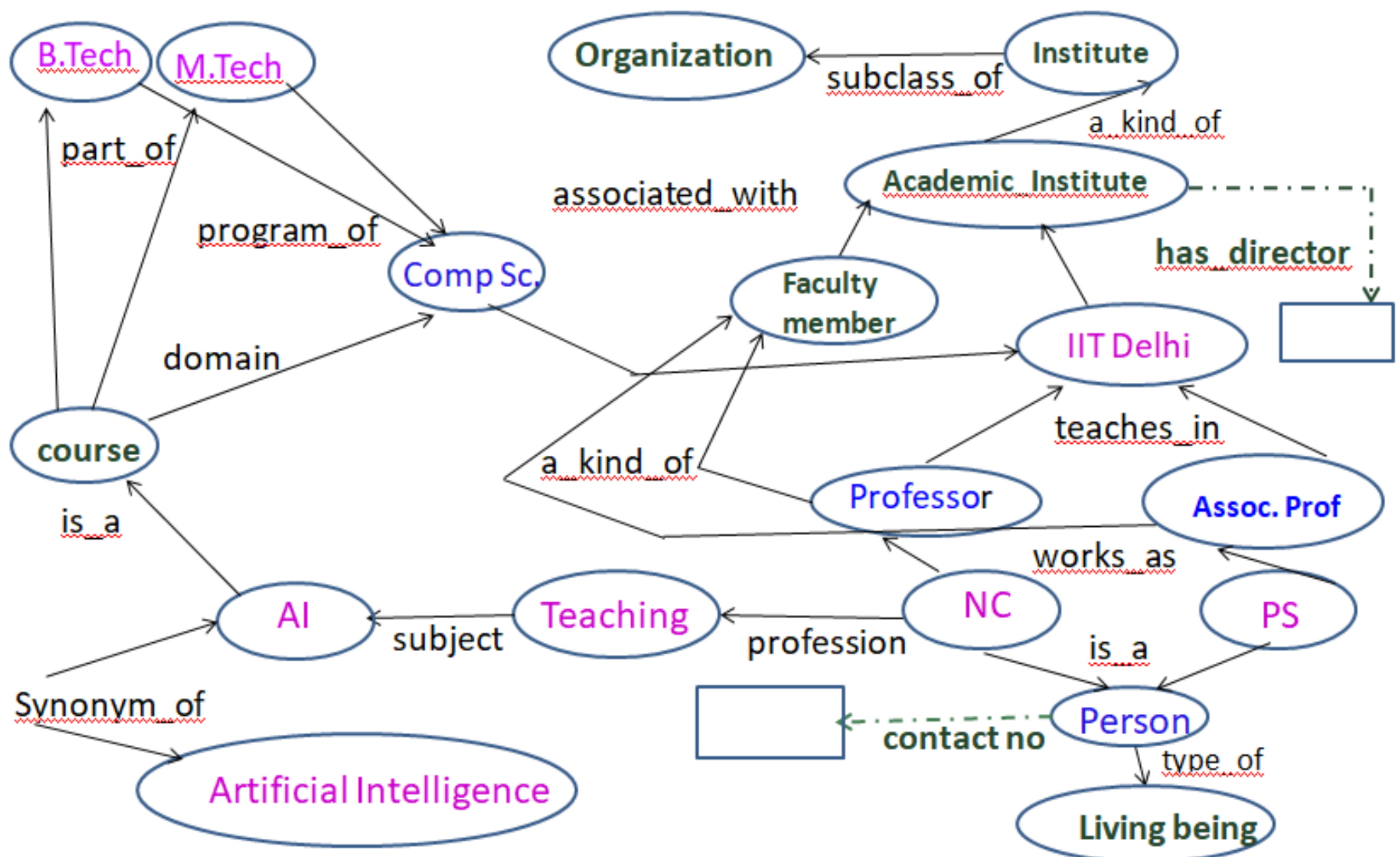
For example:

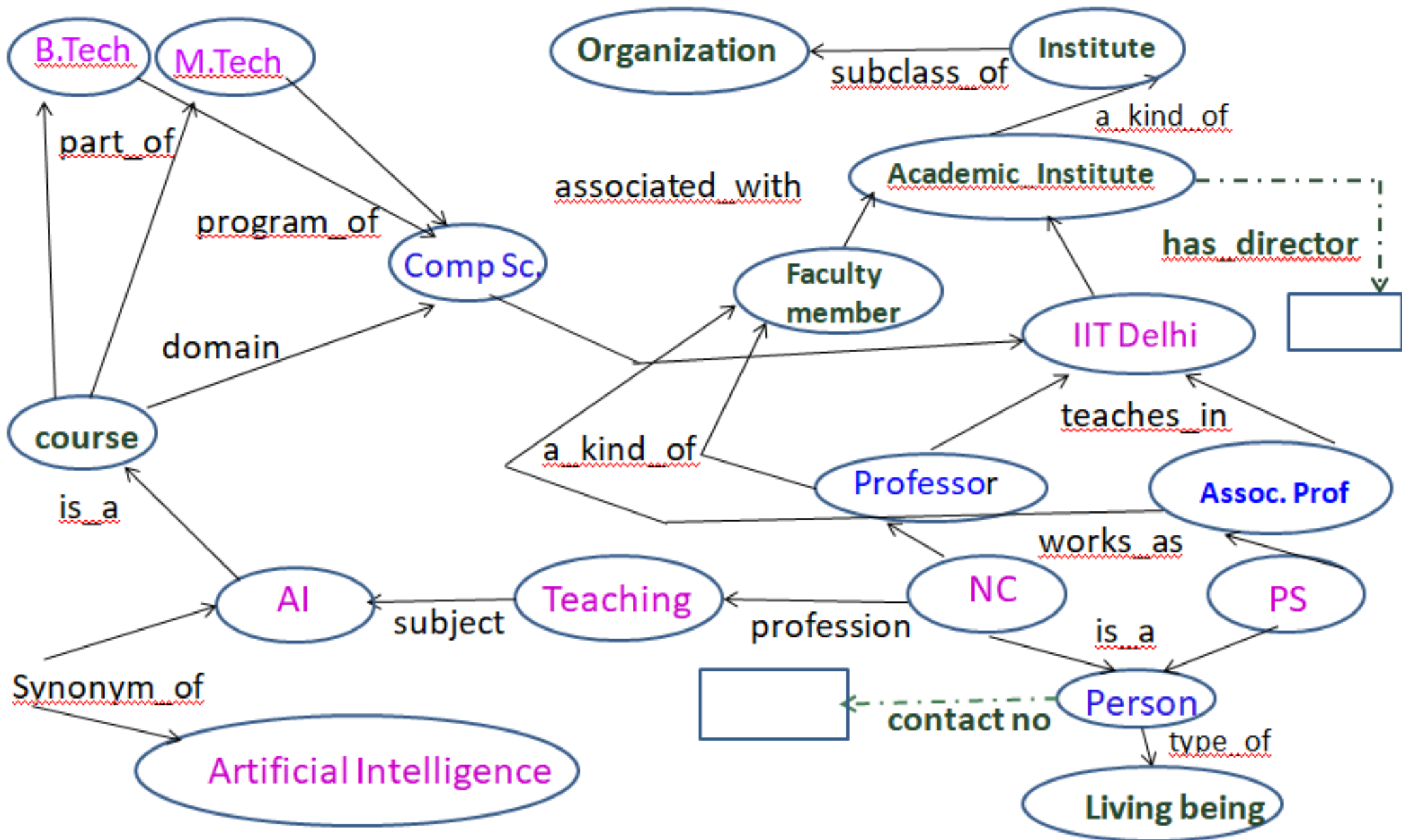
- Who is the director of IIT Delhi ?
- What is the contact number of PS ?

To answer such questions it is necessary to link some of the entities to some Global concepts for required knowledge:

- That PS is a person. And, in today's context, "contact no." is an attribute.
- That each academic institute need to have a Director.

So we may have the following diagram.





Much more information can be added depending upon the requirement.

Such a compact form of knowledge is termed as: **Ontology**

Definition: Ontology is a Formal, explicit specification of a shared conceptualization

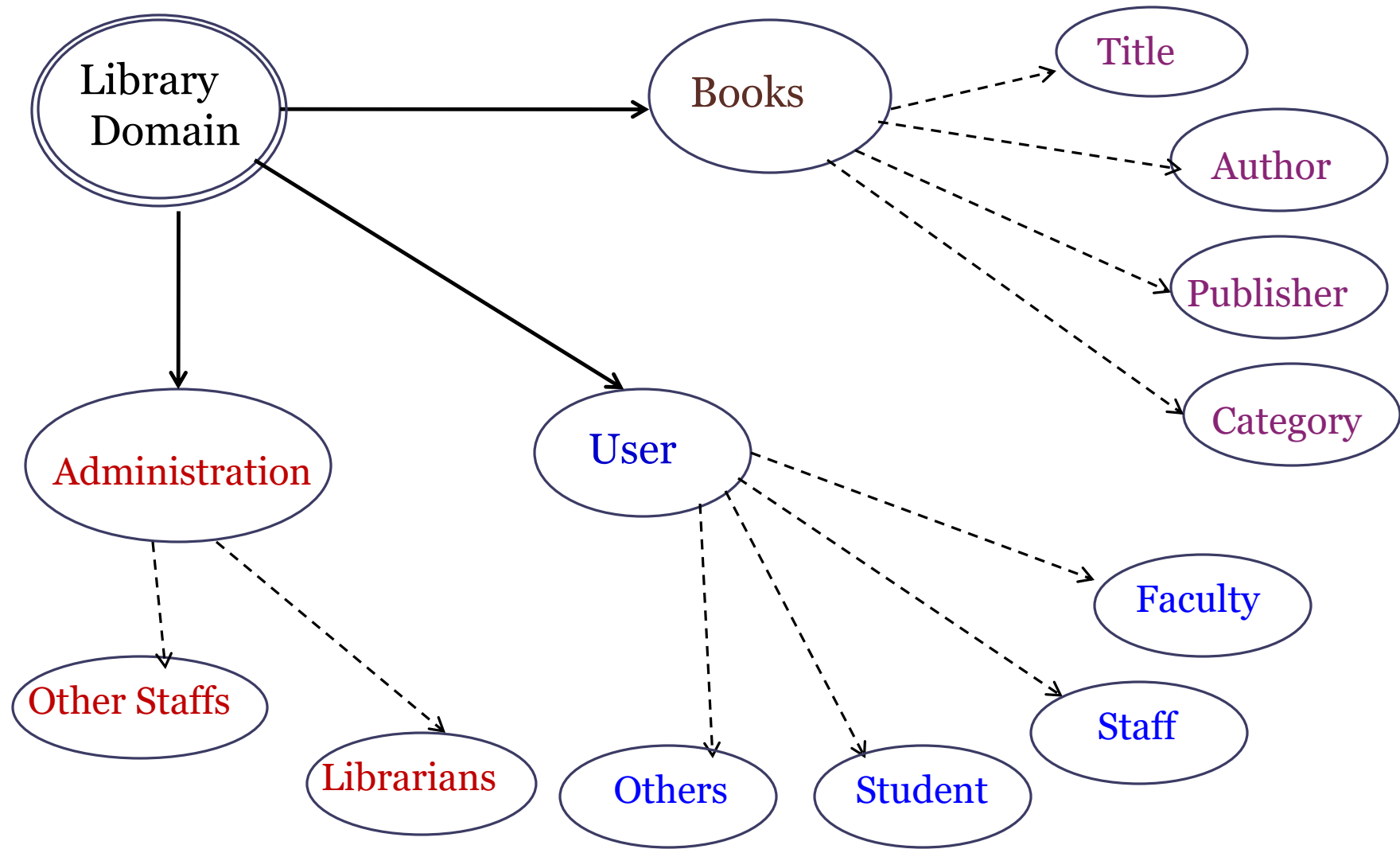
- Formally it represents knowledge as a set of concepts within a domain and also specifies the relationships among those concepts.
- It can be used to reason about the entities within that domain
- It Enables integration and interoperability of data among different communities.

Definition: Ontology is a Formal, explicit specification of a shared conceptualization

- Formally it represents knowledge as a set of concepts within a domain and also specifies the relationships among those concepts.
- It can be used to reason about the entities within that domain
- It Enables integration and interoperability of data among different communities.

In general, Ontology development process may consist of the following steps:

- Study the domain
- Find out the relevant terms or identify the concepts that one needs to include in his/her ontology
- Identify the attributes related to the term/concepts
- Identify the relationships between the concepts
- Write the constraints or rules on these concepts and relationships
- Document the ontology
- Test the ontology
- Maintenance of the ontology



How to store it inside a Computer system?

While Semantic annotation or graph based representation scheme give us intuitive idea of knowledge elicitation and knowledge acquisition – the main problem still remains:

While Semantic annotation or graph based representation scheme give us intuitive idea of knowledge elicitation and knowledge acquisition – the main problem still remains:

The problem of Knowledge Representation

While Semantic annotation or graph based representation scheme give us intuitive idea of knowledge elicitation and knowledge acquisition – the main problem still remains:

The problem of Knowledge Representation

Effective use of the knowledge is possible only if its semantic representation is efficient to access and retrieve and conducive for Inferencing.

While Semantic annotation or graph based representation scheme give us intuitive idea of knowledge elicitation and knowledge acquisition – the main problem still remains:

The problem of Knowledge Representation

Effective use of the knowledge is possible only if its semantic representation is efficient to access and retrieve and conducive for Inferencing.

A natural question arises therefore is how to represent knowledge.

While Semantic annotation or graph based representation scheme give us intuitive idea of knowledge elicitation and knowledge acquisition – the main problem still remains:

The problem of Knowledge Representation

Effective use of the knowledge is possible only if its semantic representation is efficient to access and retrieve and conducive for Inferencing.

A natural question arises therefore is how to represent knowledge.

One standard what of doing it through XML

Extensible Markup Language

Markup Language

A Markup language is a computer language that uses **tags** to define elements within a document.

It is human-readable, i.e. Markup files contain standard words, rather than typical programming syntax.

While several markup languages exist, the two most popular ones are

HTML and XML

Source: https://techterms.com/definition/markup_language

HTML (Hyper text Markup Language) is a Markup language used for creating webpages.

Its concern is the design and appearance of the page.

Illustration



test.html

new.html



```
1 <!DOCTYPE html>
2 <html lang="en">
3 <body bgcolor="#FFFFFF">
4
5 <h1>Artificial Intelligence: A Modern Approach</h1>
6 <h2>Edition 3</h2>
7 <h2>By Stuart Russel and Peter Norvig</h2>
8 <h2>ISBN-13: 978-0136042594</h2>
9 <h2>ISBN-10: 0136042597</h2>
10 </body>
```

Artificial Intelligence: A Modern Approach

Edition 3

By Stuart Russel and Peter Norvig

ISBN-13: 978-0136042594

ISBN-10: 0136042597

Stuart Russel and Peter Norvig

Artificial Intelligence: A Modern Approach

Edition 3

ISBN Numbers

- 13: 978-0136042594
- 10: 0136042597


```
1 <!DOCTYPE html>
2 <html lang="en">
3 <body bgcolor="#FFFFFF">
4 <div>
5 <h1 style="color:blue;"> Stuart Russel and Peter Norvig</h1>
6 <h1> Artificial Intelligence: A Modern Approach </h1>
7 <h2>Edition 3</h2>
8 <h2 style="font-size:20px;"><i><b>ISBN Numbers</b></i></h2>
9
10 <ul style="list-style-type:disc">
11 <p align="media-left"><li> <i><b>13: 978-0136042594</b></i></
12 <li><i><b>10: 0136042597</b></i></li>
13 |
14
15 </ul>
16
17
18 </body>
```

- The above examples make it clear the HTML tags provide only display information.
- It cannot describe the information content or the semantics
- It is not context aware
- It is not suitable for data interchange
- Hence their role in semantic interoperability is almost insignificant.

- The above examples make it clear the HTML tags provide only display information.
- It cannot describe the information content or the semantics
- It is not context aware
- It is not suitable for data interchange
- Hence their role in semantic interoperability is almost insignificant.

This leads to development of a superior markup language

Illustration:

The Computer Science book “Artificial Intelligence: A Modern Approach “ is written by Stuart Russel and Peter Norvig

Example...

```
<book name = “Artificial Intelligence: A Modern Approach”>  
  <subject name=“Computer Science”>  
    < author> Stuart Russel  and Peter Norvig</author>  
  </subject>  
</book>
```

Example...

```
<book name = “Artificial Intelligence: A Modern Approach”>  
  <subject name=“Computer Science”>  
    < author> Stuart Russel  and Peter Norvig</author>  
  </subject>  
</book>
```

```
<author name = “Stuart Russel  and Peter Norvig”>  
  < book name=“Artificial Intelligence: A Modern Approach ”>  
    <subject> Computer Science </subject>  
  </book>  
</ author >
```

Example

```
<book name = “Artificial Intelligence: A Modern Approach”>  
  <subject name=“Computer Science”>  
    < author> Stuart Russel  and Peter Norvig</author>  
  </subject>  
</book>
```

```
<author name = “Stuart Russel  and Peter Norvig”>  
  < book name=“Artificial Intelligence: A Modern Approach ”>  
    <subject> Computer Science </subject>  
  </book>  
</ author >
```

```
< subject name = “Computer Science“>  
< author> “Stuart Russel  and Peter Norvig”</author>  
  < book name=“Artificial Intelligence: A Modern Approach”> </book>  
</subject >
```

The same content may be specified with different formats without affecting semantic interpretation.

The same content may be specified with different formats without affecting semantic interpretation.

This comes in the form of XML – Extensible Markup Language

Advantages of using XML...

1. XML is independent of the platform and also independent of the underlying programming language.
2. It can be used on any system and supports any technology change.
3. No fixed set of tags – user can define their own.
4. The data stored and transported using XML can be changed without affecting the data presentation.

Source: <https://beginnersbook.com/2018/10/advantages-and-disadvantages-of-xml/>

Advantages of using XML

1. XML is independent of the platform and also independent of the underlying programming language.
2. It can be used on any system and supports any technology change.
3. No fixed set of tags – user can define their own.
4. The data stored and transported using XML can be changed without affecting the data presentation.

XML supports UNICODE. Thus XML can deal with any information written in any human language.

Source: <https://beginnersbook.com/2018/10/advantages-and-disadvantages-of-xml/>

Example XML Document

```
?xml version="1.0" encoding="ISO-8859-1"?)
```

```
<bookstore>
```

```
  <book category="COOKING">
```

```
    <title lang="en">Everyday
```

```
Italian</title>
```

```
    <author>Giada De Laurentiis</author>
```

```
    <year>2005</year>
```

```
    <price>30.00</price>
```

```
  </book>
```

```
  <book category="CHILDREN">
```

```
    <title lang="en">Harry Potter</title>
```

```
    <author>J K. Rowling</author>
```

```
    <year>2005</year>
```

```
    <price>29.99</price>
```

```
  </book>
```

```
</bookstore> book category="WEB">
```

```
  <title lang="en">XQuery Kick Start</title>
```

```
  <author>James McGovern</author>
```

```
  <author>Per Bothner</author>
```

```
  <author>Kurt Cagle</author>
```

```
  <author>James Linn</author>
```

```
  <year>2003</year>
```

```
  <price>49.99</price>
```

```
</book>
```

```
<book category="WEB">
```

```
  <title lang="en">Learning XML</title>
```

```
  <author>Erik T. Ray</author>
```

```
  <year>2003</year>
```

```
  <price>39.95</price>
```

```
</book>
```

```
</bookstore>
```

It does not mean that we cannot add presentation styles with XML.

XSLT is used for the same.

XSLT...

XSLT (eXtensible Stylesheet Language Transformations) is a language for transforming XML documents into other XML documents, or other formats such as HTML for web pages, plain text or XSL Formatting Objects, which may subsequently be converted to other formats, such as PDF, PostScript and PNG.

The original document is not changed; rather, a new document is created based on the content of an existing one.

Typically, input documents are XML files, but anything from which the processor can build an XQuery and XPath Data Model can be used, such as relational database tables or geographical information systems

Source: <https://en.wikipedia.org/wiki/XSLT>

XSLT

Suggested Reading:

https://www.w3schools.com/xml/xsl_intro.asp

Before I stop I give you some terms that are important in this domain:

Name Space

- An abstract container that holds a logical grouping of unique identifiers or symbols
- An identifier defined in a namespace is associated only with that namespace.
- The same identifier can be independently defined in multiple namespaces.
- Meaning associated with an identifier defined in one namespace may not have the same meaning in another namespace.

UNICODE : Standard for the consistent encoding, representation and handling of text (1987).

XML Schema : A language used to define the structure of specific XML document.

RDF : Resource Description Framework

A flexible language capable of describing all sorts of information and Metadata. Used as a general method for conceptual description of information. Typically in the form of triplets <Entity Attribute Value>.



THANK YOU

