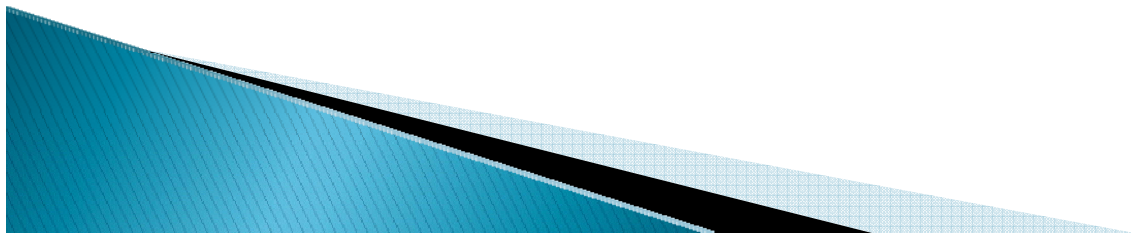# Classificatory Ontologies

**ARD Prasad**

**Devika P. Madalli**

Documentation Research and Training Center,(DRTC),
Indian Statistical Institute, Bangalore, India

# Semantic Web & Metadata

- *Metadata* a key concept of the envisioned Semantic Web

- Metadata is not a new concept for librarians.

- Practiced in libraries since the days of Cutter (and before that too!)

International UDC Seminar, The Hague, Oct 2009

# DLs & Metadata

- Various types of metadata, e.g. Descriptive, Administrative, Structural & Subject are used to provide a varied range of information.

- *Subject Metadata* is presented through some Knowledge Organization Systems, e.g. classification schemes, thesauri, ontologies, etc.

# Focus on Subject Approach

- Our focus is on "Subject Approach to Information" according to Ranganathan's facet analysis.

International UDC Seminar, The Hague,  Oct 2009

# Classificatory Ontologies

- Our approach suggests using classificatory ontologies in DLs for KO
- Further, ontologies to be based on Faceted Classification

International UDC Seminar, The Hague, Oct 2009

# Why Faceted Classification?

International UDC Seminar, The Hague,  Oct 2009

# Establishing the purpose

- Some <u>questions</u> which the knowledge stored in an ontology should answer

  - I want to know about the homepage of Tim Berners- Lee
  - I want to how the persons with whom he has written papers

International UDC Seminar, The Hague,  Oct 2009

# Queries on a discourse

**The discourse**

Semantic Web is started with a vision of the inventor of the WWW, Tim Berners-Lee.

He is also the director of World Wide Web Consortium (W3C).

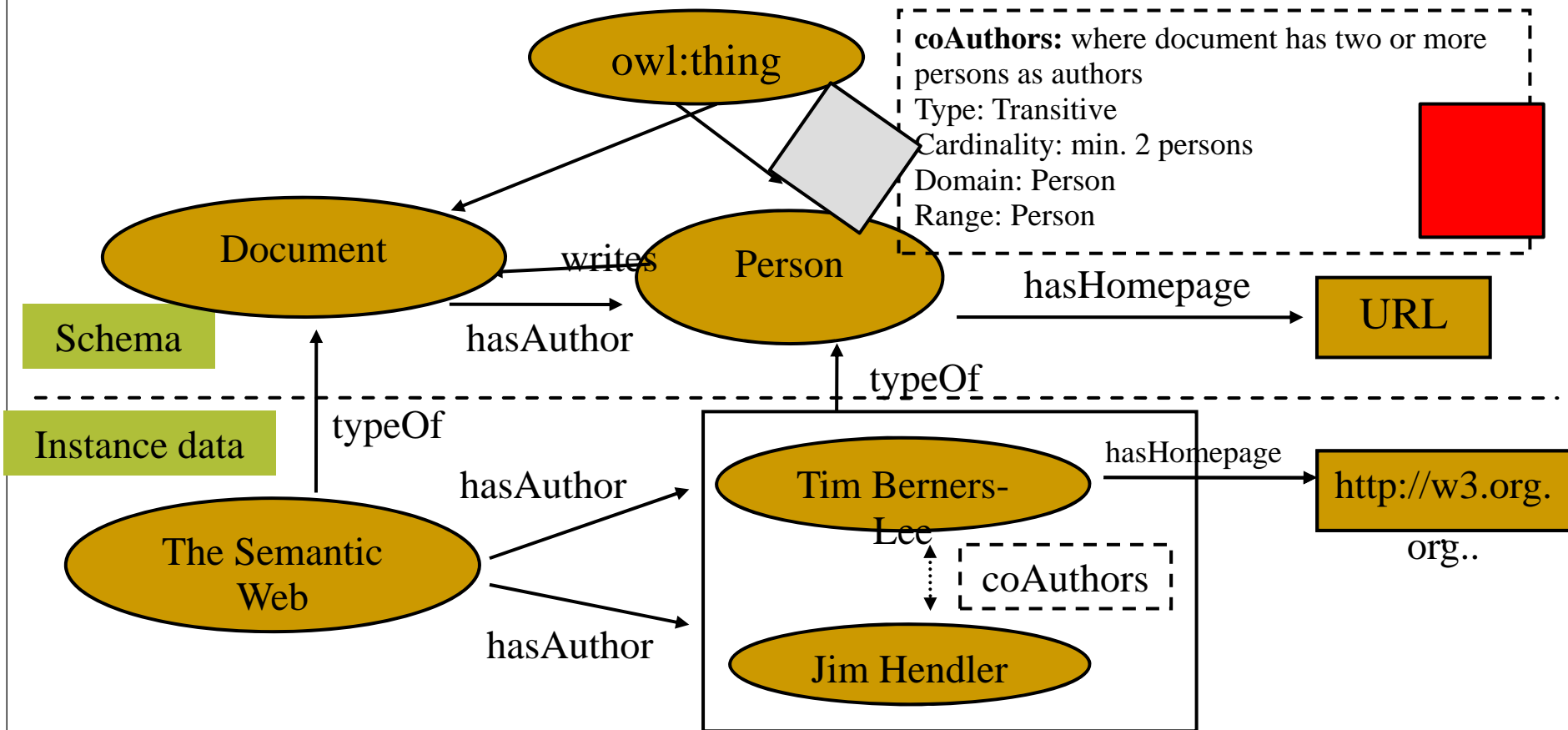All his papers on SW can be accessed through his homepage (http://w3.org/people/tbl).

One of the initial papers on SW is 'The Semantic Web' written in collaboration with Jim Hendler and Ora Lassila in 2001.

*What is the URL of Tim Berners-Lee's homepage*

*Who are the co-authors of Tim Berners-Lee*

International UDC Seminar, The Hague, Oct 2009

# Ontology: Generic Model

owl:thing

coAuthors: where document has two or more persons as authors
Type: Transitive
Cardinality: min. 2 persons
Domain: Person
Range: Person

Document

writes

Person

hasHomepage

URL

Schema

hasAuthor

typeOf

Instance data

typeOf

hasAuthor

Tim Berners-Lee

hasHomepage

http://w3.org.
org..

The Semantic Web

coAuthors

hasAuthor

Jim Hendler

# Query on 'subject' of documents

*What are the documents on "Effect of nitrogen fertilizers on rice plants."*

- Many types of fertilizers: but only nitrogen based.

- Effect of any part of the plant

- Effect on different varieties of rice plants

- Effect of any type of nitrogen fertilizers on any part of plant
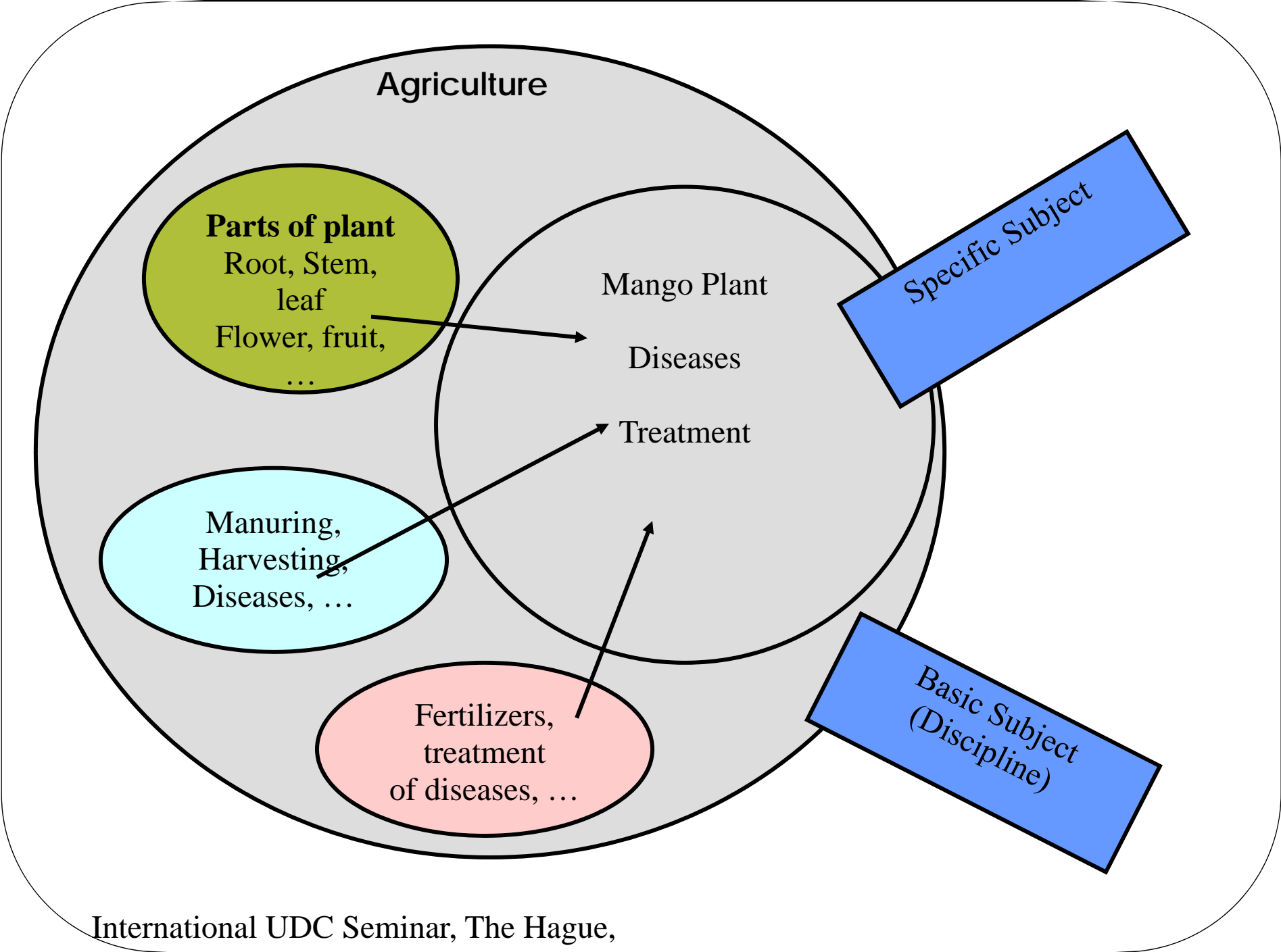
- Different methods of manuring

# What is Subject? (in general)

- The subject is a unit fact *

- It summarizes in itself the message what a particular body of information is about *

  - 'Parts of plants, and their utility', 'manuring, harvesting, production of crops', methods, things used for manuring, harvesting, etc'

\* G. Bhattacharyya. Role of classification and indexing in information retrieval. DRTC Annual Seminar (20), 1983. Paper AA

# What is 'Subject' of a document

- Called as 'Specific Subject" By Ranganathan

- *"The specific subject* (of the document) *is that division of knowledge whose extension and intention are equal to those of its thought content"* (Ranganathan, 1960, pp. 4).

Agriculture

**Parts of plant**
Root, Stem, leaf
Flower, fruit, …

Mango Plant

Diseases

Treatment

Specific Subject

Manuring, Harvesting, Diseases, …

Fertilizers, treatment of diseases, …

Basic Subject (Discipline)

International UDC Seminar, The Hague, Oct 2009

Hence,

Subject description = 'Thought content' of the document

# Why Faceted Approach to Subject Classification/Indexing

**Contextual Retrieval**

- What are the available contexts in which the term **'Virus'** is appearing in the Knowledgebase

- Virus [propertyOfDisease]> Rice [TypeOf]> Plant [Entity] > Agriculture [Discipline]  → context

- Virus [propertyOfDisease]> Body [partOf]> Human[Entity] > Medicine [Discipline]  → context

# Why Faceted Approach to Subject Classification/Indexing

**Inference based Retrieval**

- Non-faceted approach of subject description

  <dc:subject>Agriculture</dc:subject>

  <dc:subject>diseases</dc:subject>

  <dc:subject>Rice</dc:subject>

- How shall the machine determine documents on from the above subject description
  - "rice crop destroyed by virus"

# What are facets?

- **Facet** is the set of classes got by applying characteristics successively

- Many such facets constitutes a basic subject

**Agriculture:**

| (By Parts of Plant) | (By Processes) | (By Property of Action) |
|---|---|---|
| ■Root | ■Manuring | ■Substance of Manuring |
| ■Stem | ■Harvesting | ■Material for Harvesting |
| ■Leaf | ■Propogation | |

| Part Facet | Process Facet | Property Facet |
|---|---|---|

Parts of plant
Root, Stem, leaf
Flower, fruit, … ← Entity

Processes
Manuring,
Harvesting,
Diseases, … ← Action

Property
Fertilizers,
treatment
of diseases, … ← Property

Elementary
Categories

Agriculture
(Discipline)
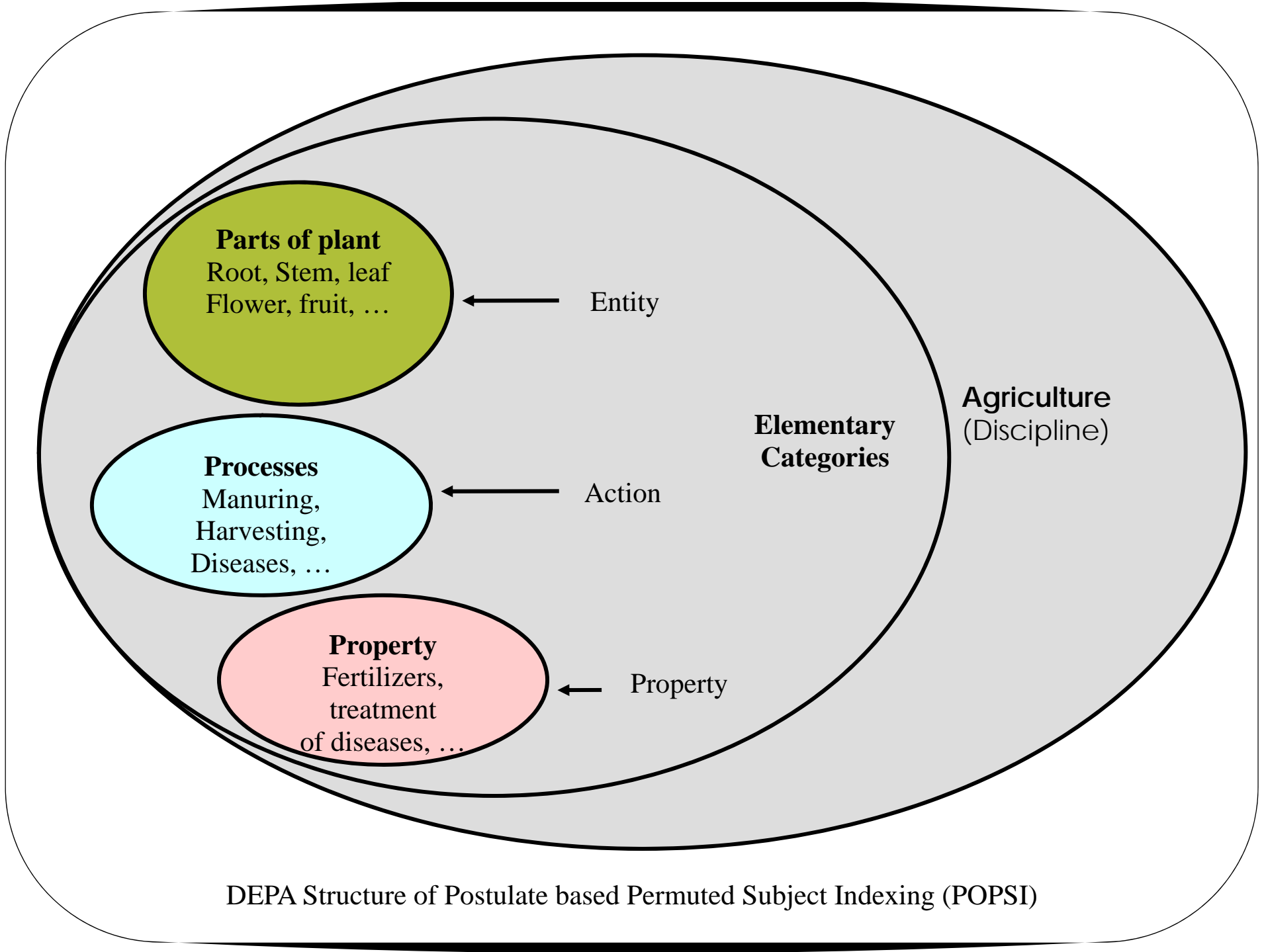
DEPA Structure of Postulate based Permuted Subject Indexing (POPSI)

# To know the "Aboutness" of Documents
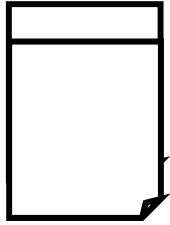
Subject Analysis of document

**Title:**

Report on production of rice using nitrogen based fertilizers

**Expressive Title:**

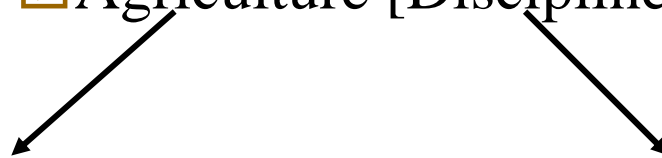In Agriculture, rice production using nitrogen based fertilizers

**Concept Identification:**

In Agriculture, plant, rice, production, manuring, manures, fertilizers, nitrogen fertilizers

# Document Aboutness …

Facet Group (Discipline):

☑Agriculture [Discipline]

Facet Group (Entity): ——→ Facet Group (Action):
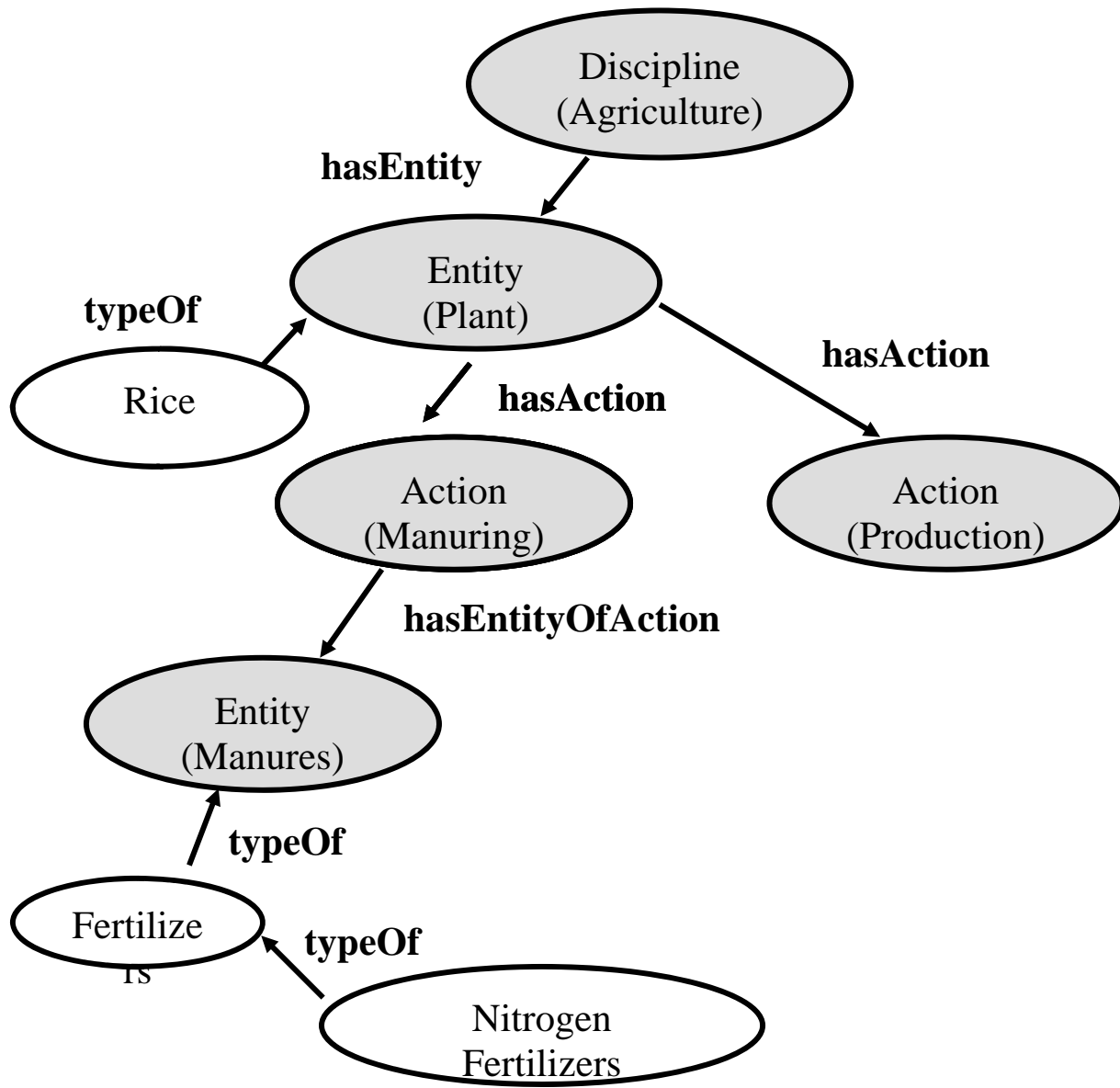
☑Plant [Entity]

☑Production [Action]

☐Rice [typeOf]

☑Manuring [Action]

☐Manures [entityOfAction]
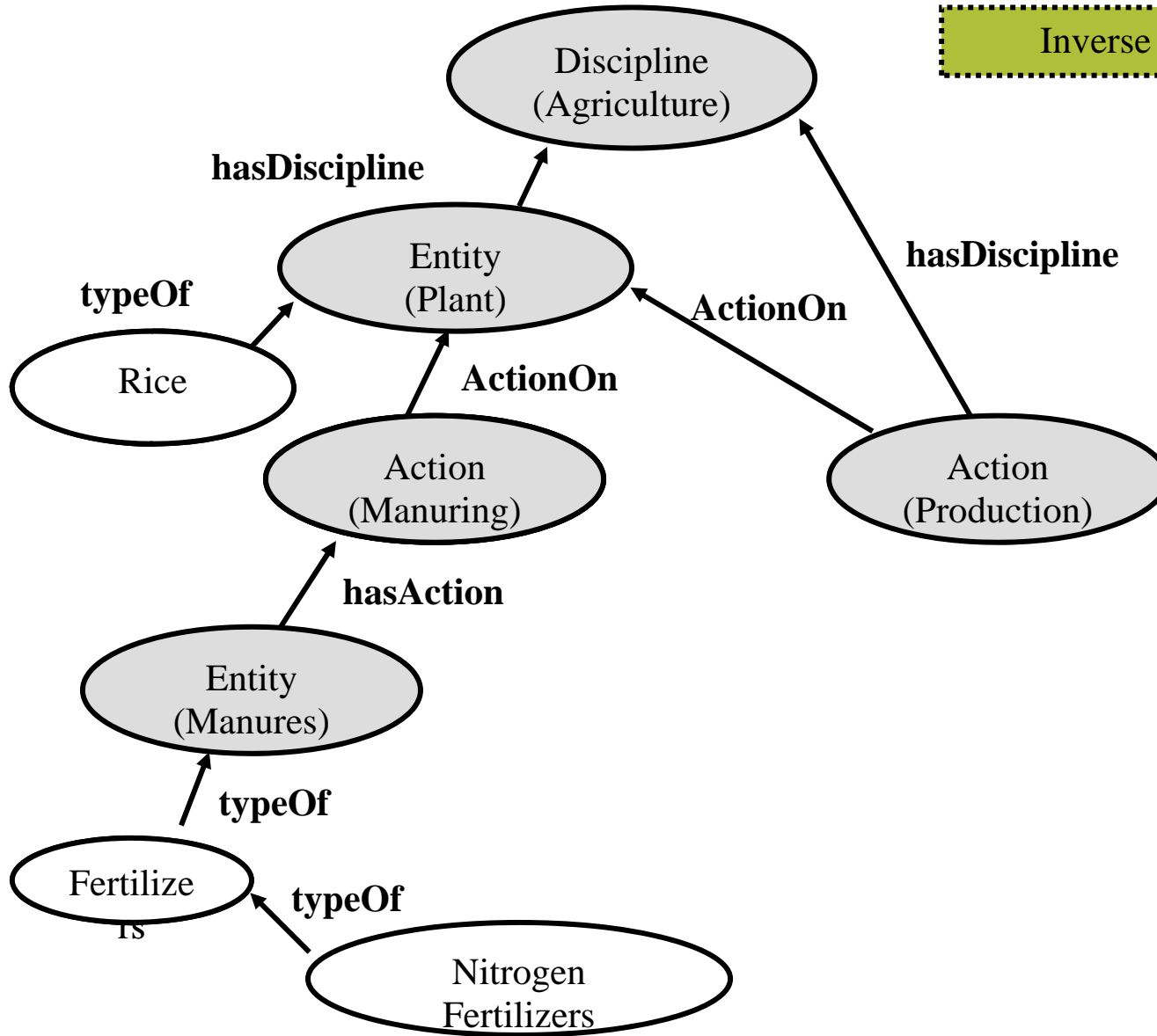
▪ Fertilizers [typeOf]

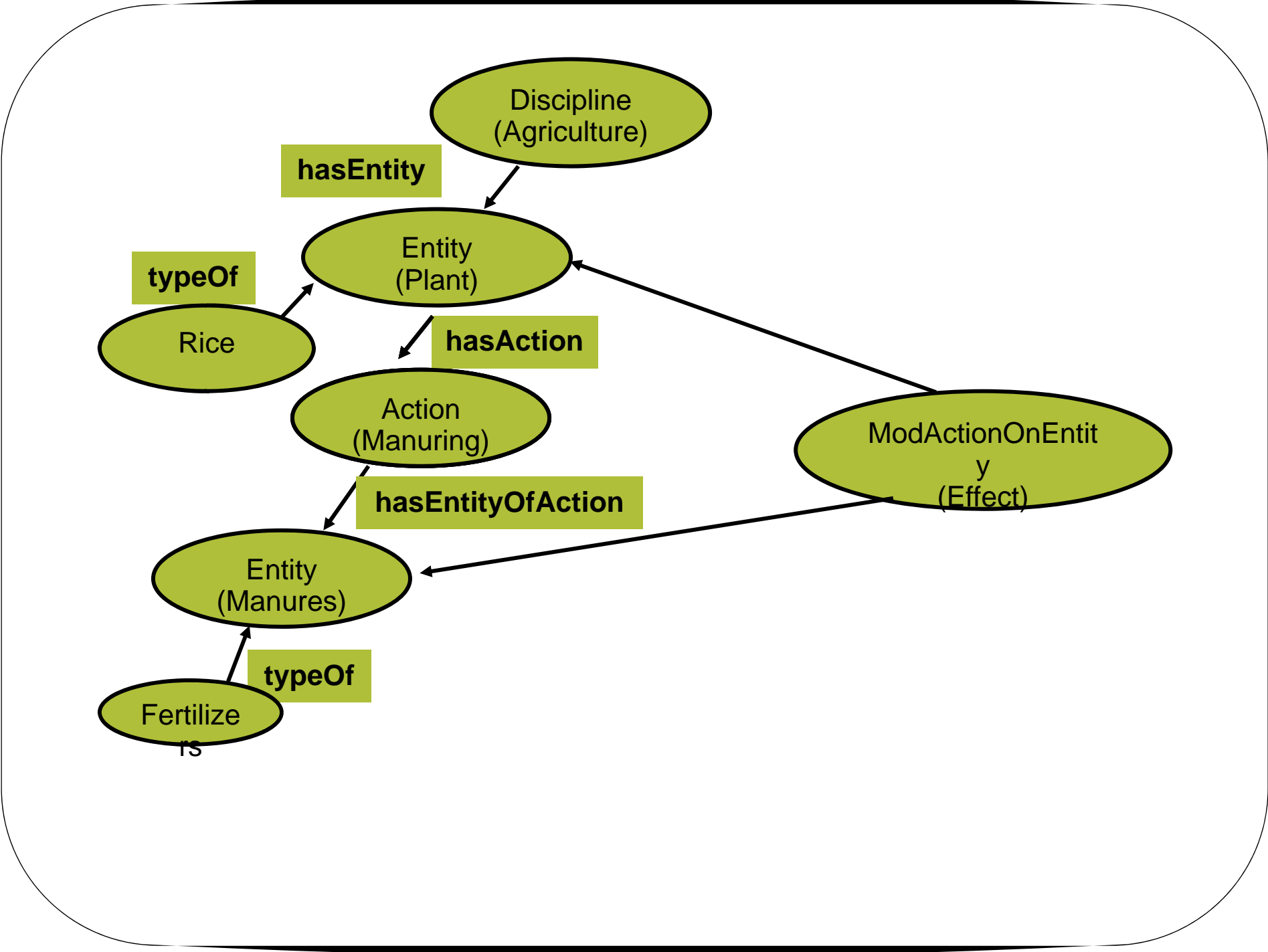▪ Nitrogen fertilizers [typeOf]

Document Knowledge

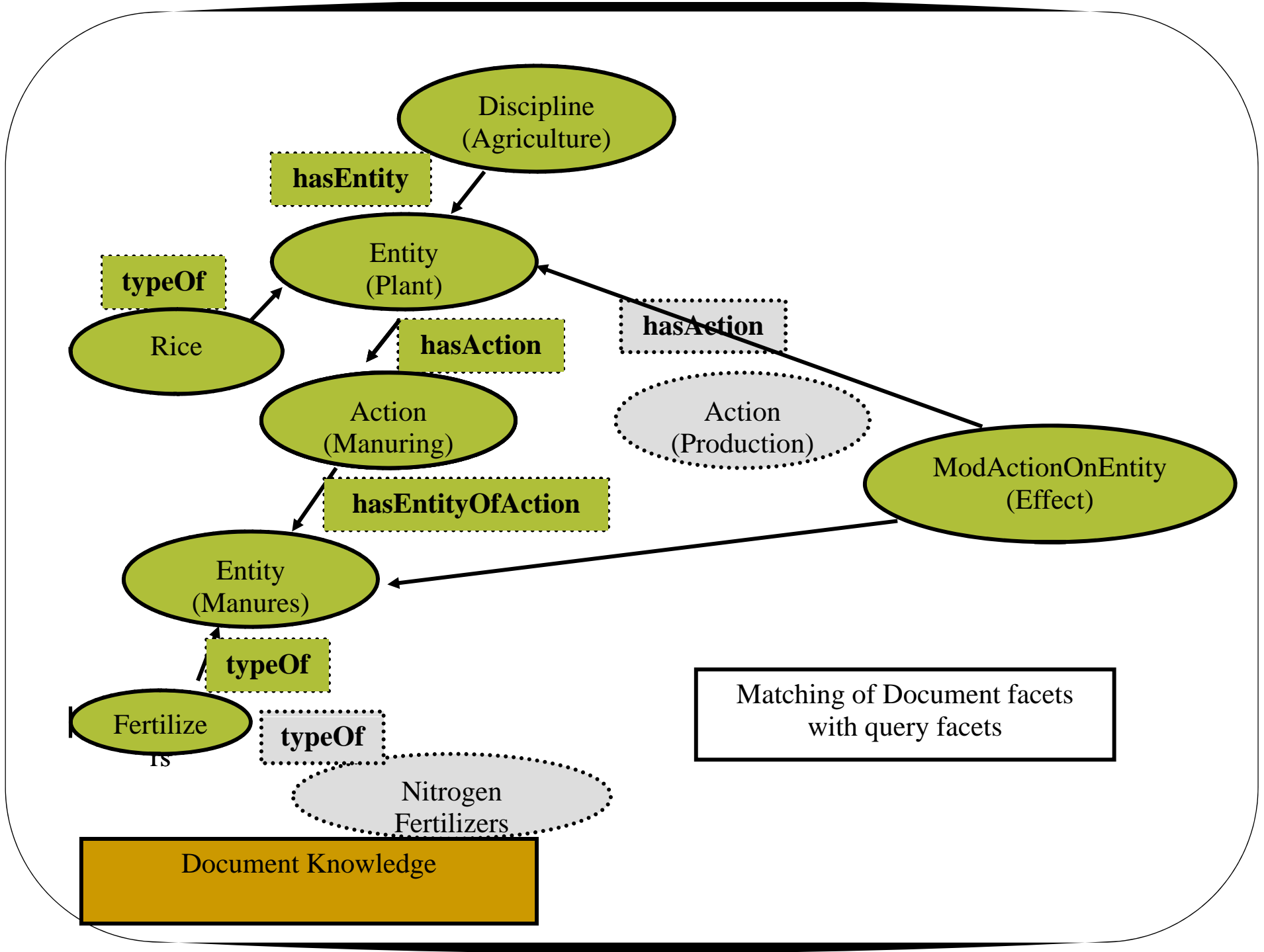# Query Analysis & Document Matching

*Effect of manures on the rice plants*

- Query Analysis:

  - Plant: Entity
  - Rice: typeOfEntity
  - Manuring: Action
  - Manures: EntityOfAction
  - Effect: ModifierActionOnEntity

Discipline (Agriculture)

hasEntity

typeOf

Entity (Plant)

Rice

hasAction

hasAction

Action (Manuring)

Action (Production)

ModActionOnEntity (Effect)

hasEntityOfAction

Entity (Manures)

typeOf

Fertilizers

typeOf

Nitrogen Fertilizers

Matching of Document facets with query facets

Document Knowledge

# Classificatory Ontologies & SKOS

- SKOS intended to be used for three important information retrieval applications viz.
  - Vocabulary Development
  - Indexing and
  - Searching

# SKOS-Good for Thesauri Representation

- SKOS can clearly represent properties of concepts, relationships among or between the concepts- both hierarchical and associative

- Hierarchical relations can be displayed using: skos:hasTopConcept, skos:broader, skos:narrower etc.

- While associative relations can be represented using skos:related

# But not best option for classification schemes..contd

- The skos:notation element does not record how a classification notation is built, which restrict the further manipulation of the class numbers in an machine assisted environment

- Though SKOS provides skos:hasTopConcept for classes in the main schedule, no provision is available to represent auxiliary tables without semantic loss[1]

1. Zeng, M. L., Fan, W. Lin, X. (2008), "SKOS for an Integrated Vocabulary Structure" in Metadata for Semantic and Social Applications, Proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22-26 September 2008, Universitätsverlag Göttingen, Berlin, pp. 200-201.

# Modeling CC in SKOS

- Directly SKOS does not support the requisites for representing faceted classification schemes

- Unlike enumerative schemes, structure of faceted schemes mainly consists of isolates distributed over fundamental categories and its rounds and levels

- Furthermore, analytico-synthetic structure provides provisions for combining the various basic subjects, isolates, etc. to represent more complex subjects

**A report on eliminating drug use among urban youth in Northern India in the year 2008**

**Facets**

      Basic subject: Sociology
      Personality: Youth (by age)
             Urban (by Residence)
      Energy: Drug habits (round1)
              Treatment (round2)
      Space Isolate: North India
      Time Isolate: Year 2008
      Common Isolate: Report

Its clear that for representing it in SKOS we need proper labeling for Basic subject, personality (two isolates superimposed), two rounds of energy, and labels for space, time and other common isolates

# Conclusion

- Need to develop extensions for SKOS so that complex structure of faceted systems can be represented
- Or development of a KOS suitable for faceted structure? (Facet-KOS)

**Thank you!**