Creating Ontology from Persian Thesauri

Molouk Sadat Hosseini Beheshti

Beheshti@irandoc.ac.ir

Faculty Member and Researcher

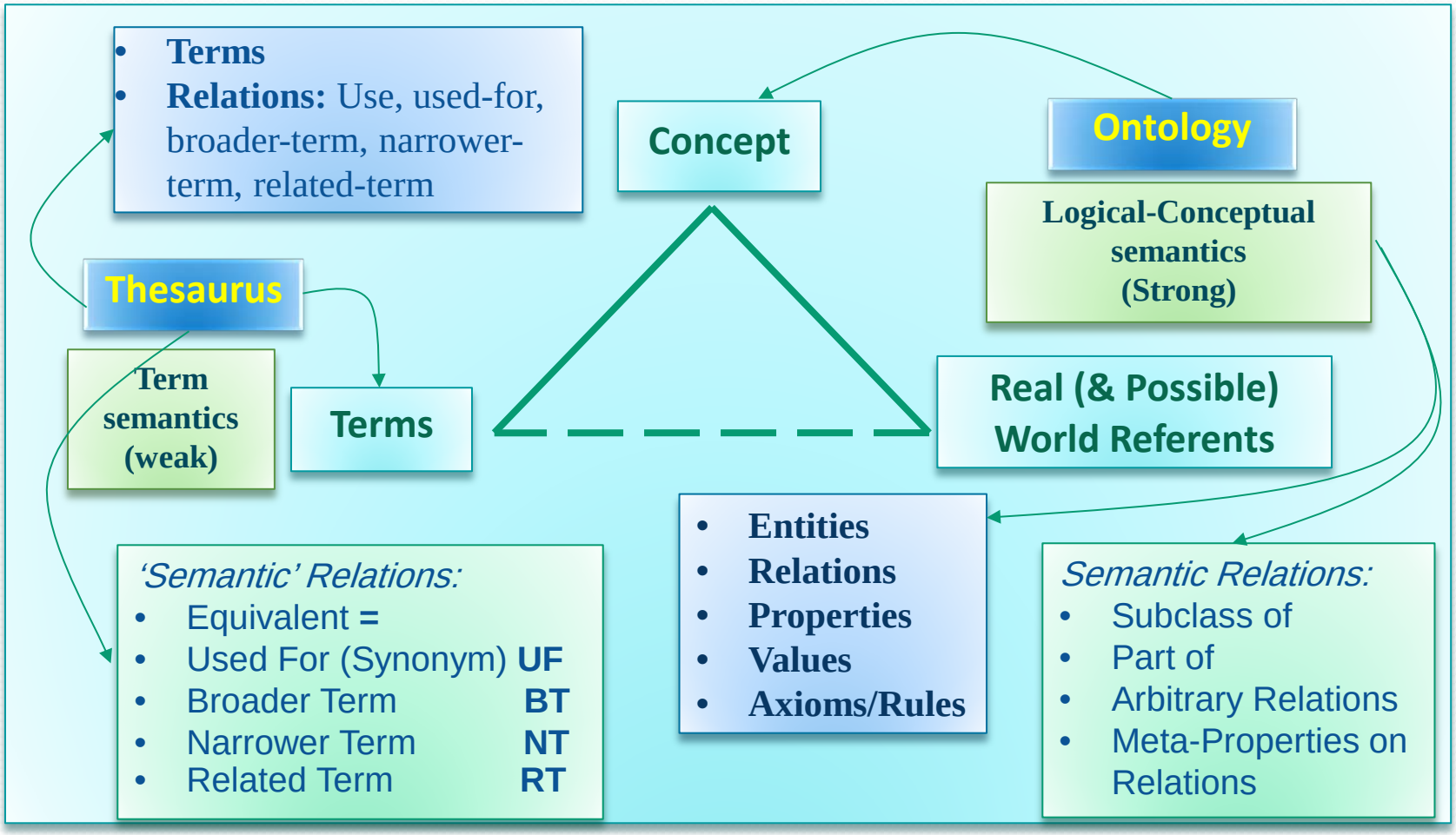
Iranian Research Institute for Information Science and Technology (IRANDOC)

Fatemeh Ejei

f_ejei@alum.sharif.edu

Sharif University of Technology

Thesaurus versus Ontology



"Thesaurus versus Ontology" Daconta, Obrst, and Smith (2003)

Creating ontology from thesaurus

- Thesaurus contains semantic information, which can be transformed into ontology
- The explicit semantic structure of thesaurus facilitates extracting term and relationships needed for ontology construction
- Hierarchical relationship of thesaurus (BT/NT) can be used to express super/subclass-type relationships of ontology

Integrating Thesauri

- We use bilingual (Persian/English) thesauri of basic sciences:
  - Chemistry
  - Physics
  - Biology
  - Geology
  - Mathematics
- A collection of tens of thousands of terms that are produced in different times and by different experts
- Thesauri data should be synchronized before using it for ontology construction

Synchronizing Concepts

- **Thesaurus Synchronizer** was developed using *Thesaurus Builder* to illustrate differences between matched cases of two thesauri
- The differences between thesauri were examined based on ISO 25964 standard
- Domain experts synchronized each two thesaurus semi-automatically
- All thesauri were merged after synchronization

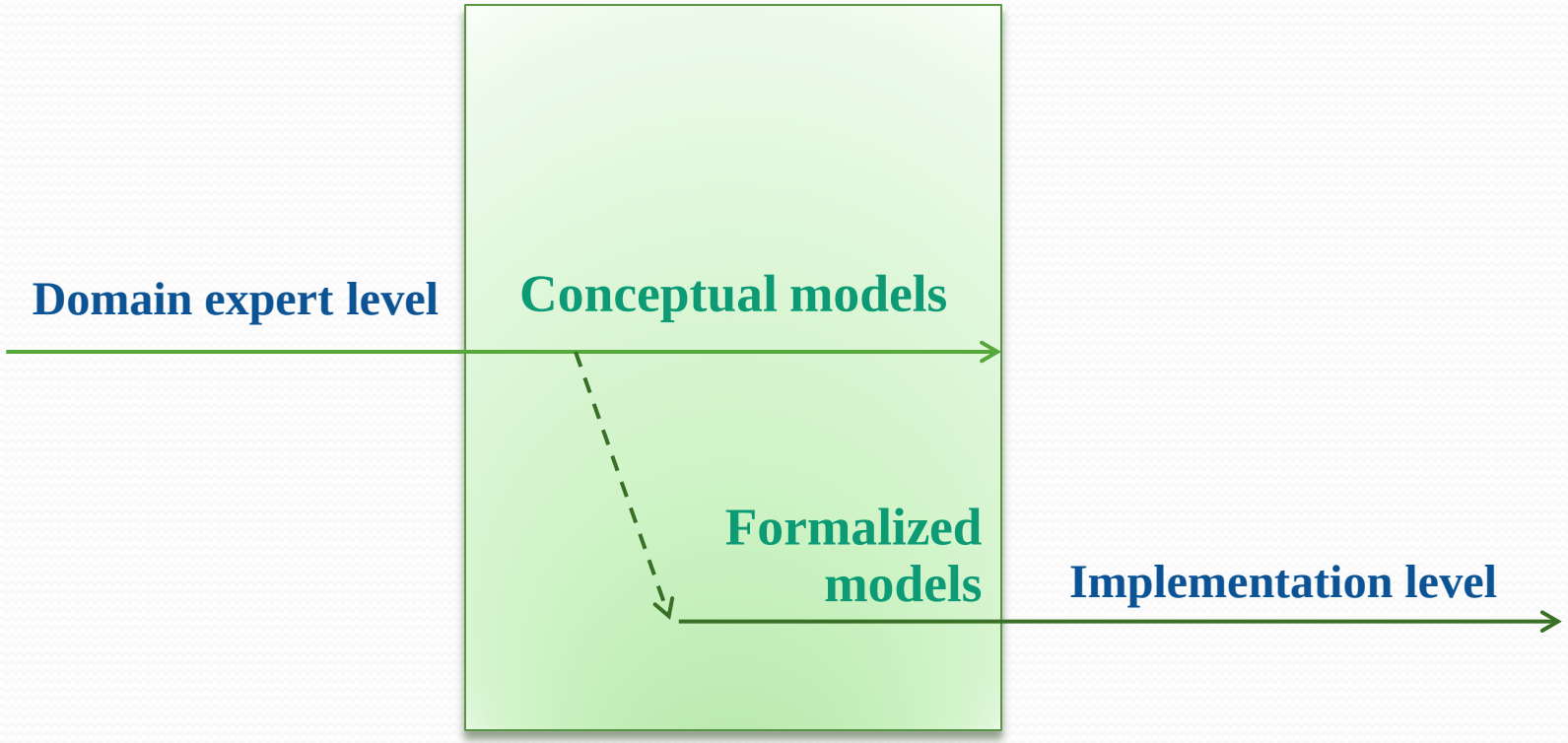
Synchronizing Concepts (cont.)

- Issues examined within each two thesauri:
  1. Differences in transcription of the same concept
  2. Differences in *narrower terms* of the same concept
  3. Differences in *non-preferred* terms of the same concept
  4. Differences between the translations of the same term in a specific language
  5. Differences in *related terms* of the same concept

Synchronizing Concepts (cont.)

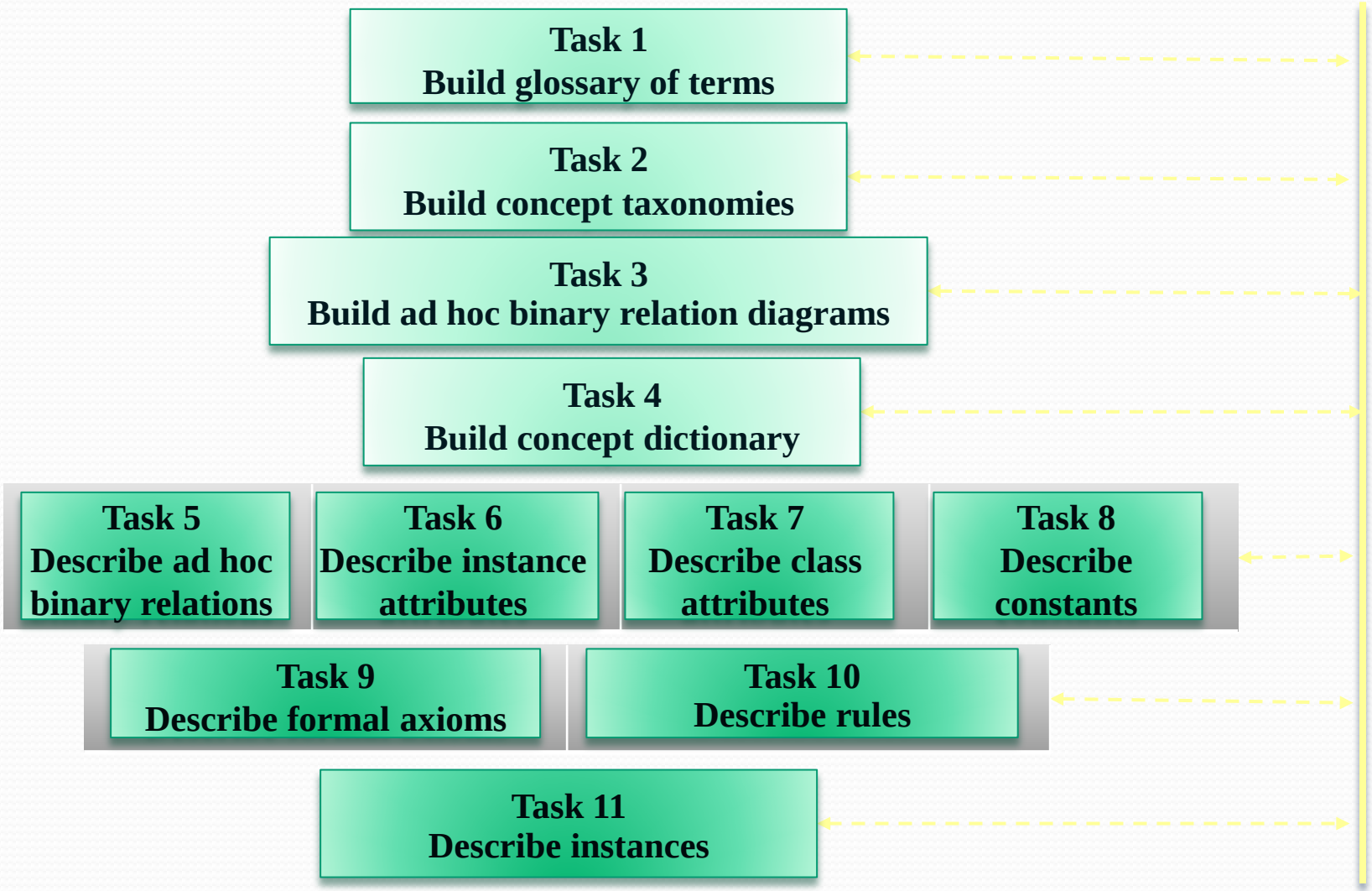
- Issues examined within each two thesauri (continued):
  6. Lack of a *related term* for a concept in one thesaurus
  7. Using the same translation for two different terms in a particular language
  8. Different selection of a *preferred term* for one concept
  9. Different relationship type between two concepts
  10. Infinite loop between concepts (conceptual network)
  11. Different concepts related to the same term (polysemy)

Ontology Development



"Essential process model in ontology development" Gómez-Pérez (2004)

Ontology Conceptualization in METHONTOLOGY



"Tasks of the conceptualization activity according to METHONTOLOGY" Gómez-Pérez (2004)

Ontology Conceptualization

- During conceptualization activity, following tasks were performed:
  1. Building glossary of terms
  2. Building concept taxonomies
  3. Building ad hoc binary relation diagrams
  4. Building concept dictionary

Building Glossary of Terms

- The glossary of terms, identifies the set of terms to be included on the ontology.
  - Including all the relevant terms of the domain:
    - Concepts
    - Instances
    - Relation between concepts
    - …
- Building based on terms and relations in thesaurus
  - Inductive method
  - Deductive method

Building Concept Taxonomies

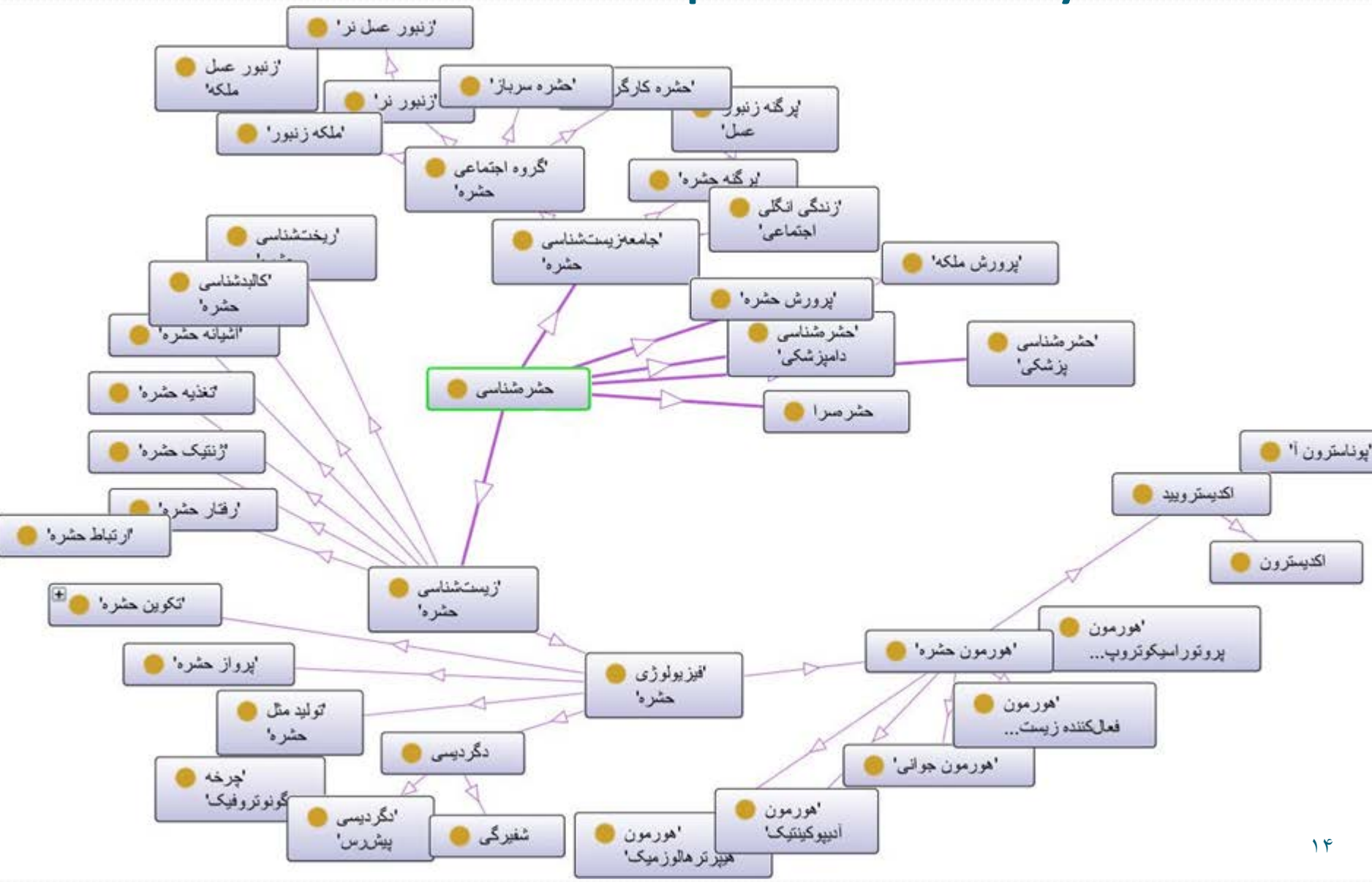
- Concept taxonomies are formed based on the taxonomic relations in thesaurus
  - Each **descriptor** (preferred term) designates a **concept**
  - **BT/NT** is converted to **super/subclass**-type relationship

X <Narrower Term> Y → X <subClassof> Y

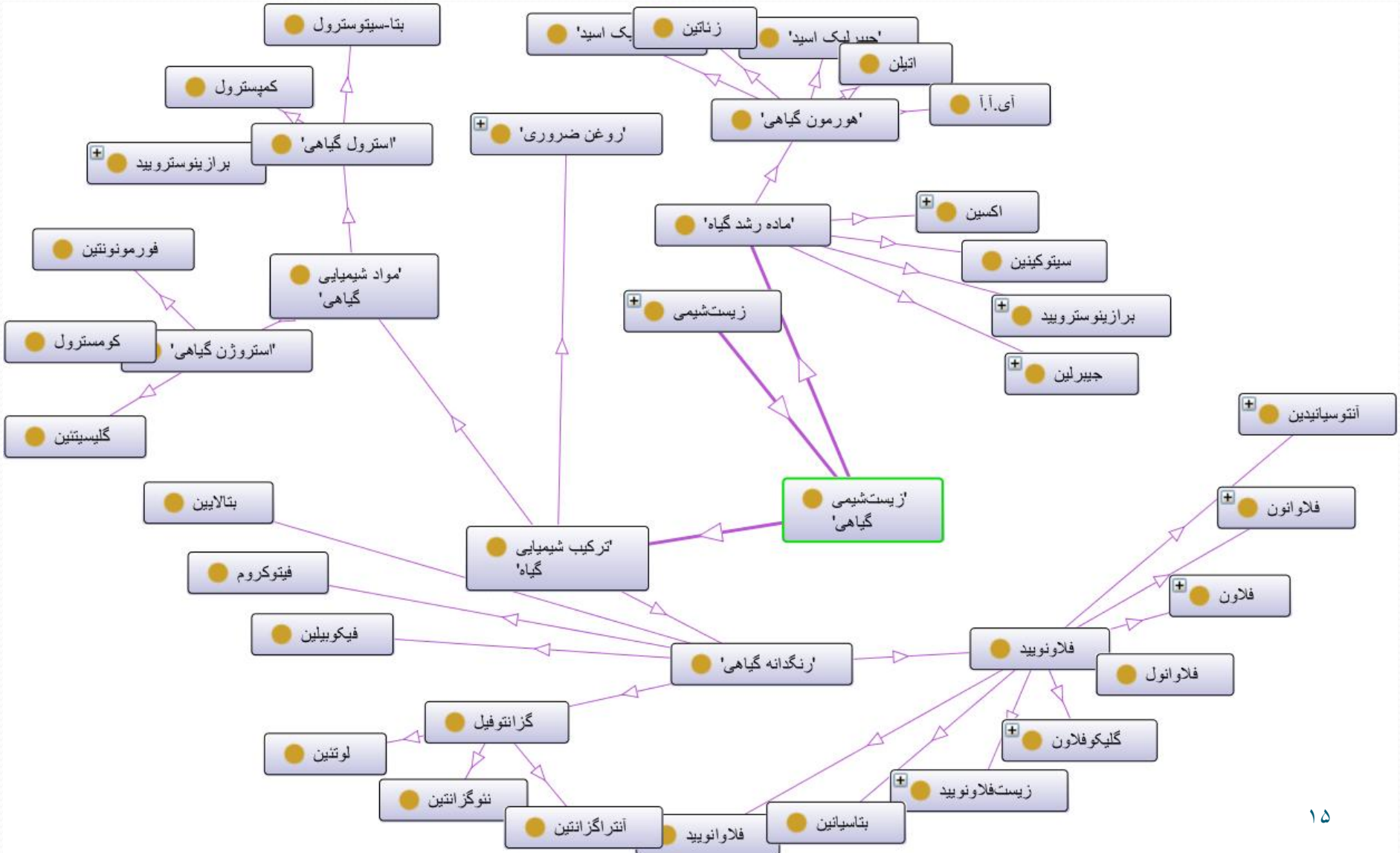
Building Ad hoc Binary Relation Diagrams

- Identifying ad hoc relationships between concepts of the ontology
- Mapping the relationships between terms in thesaurus into semantic relationships between corresponding concepts in ontology
- Each relationship labeled with its corresponding relationship type (BT, NT, RT) in thesaurus.

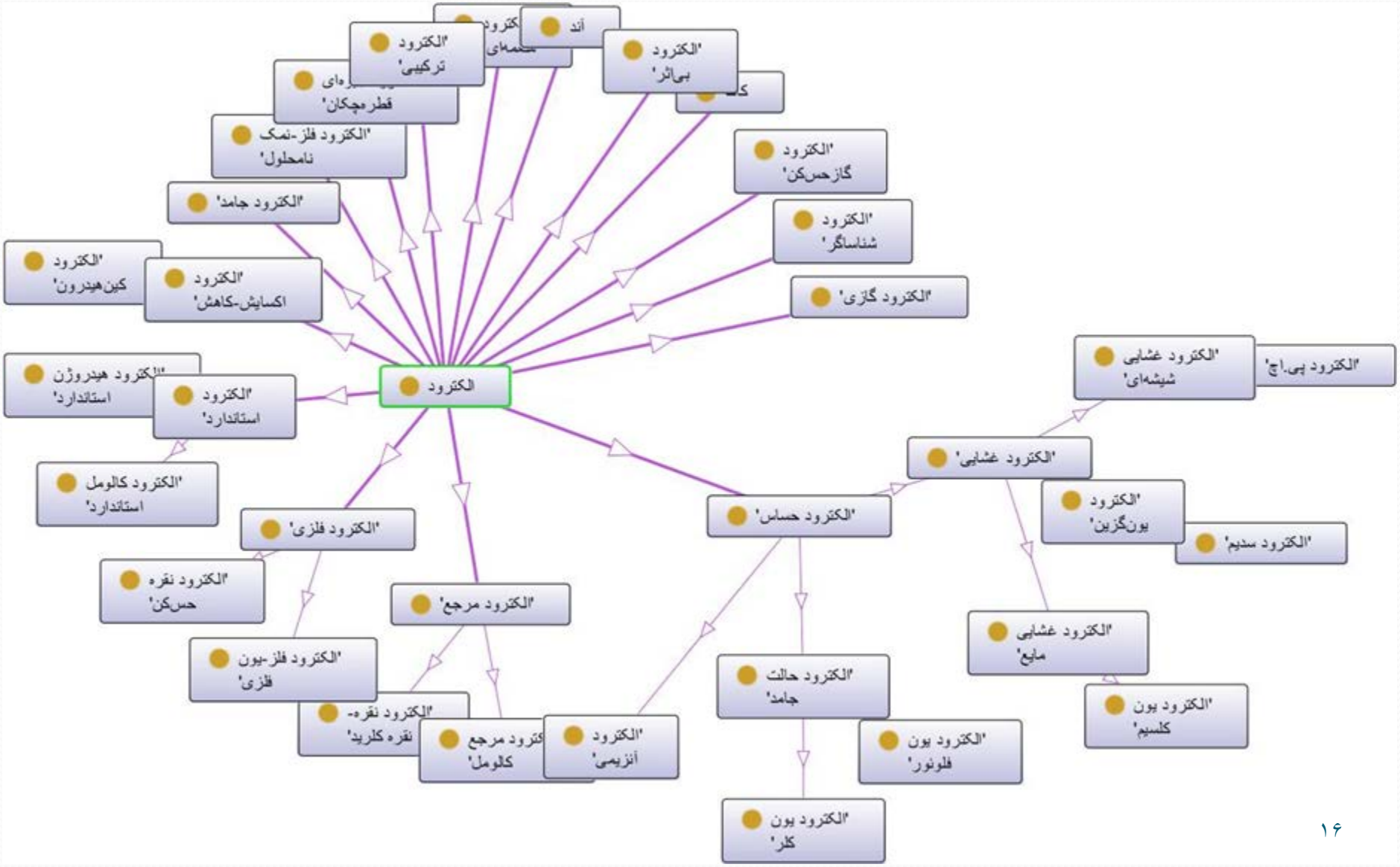
Part of Concept Taxonomy 1



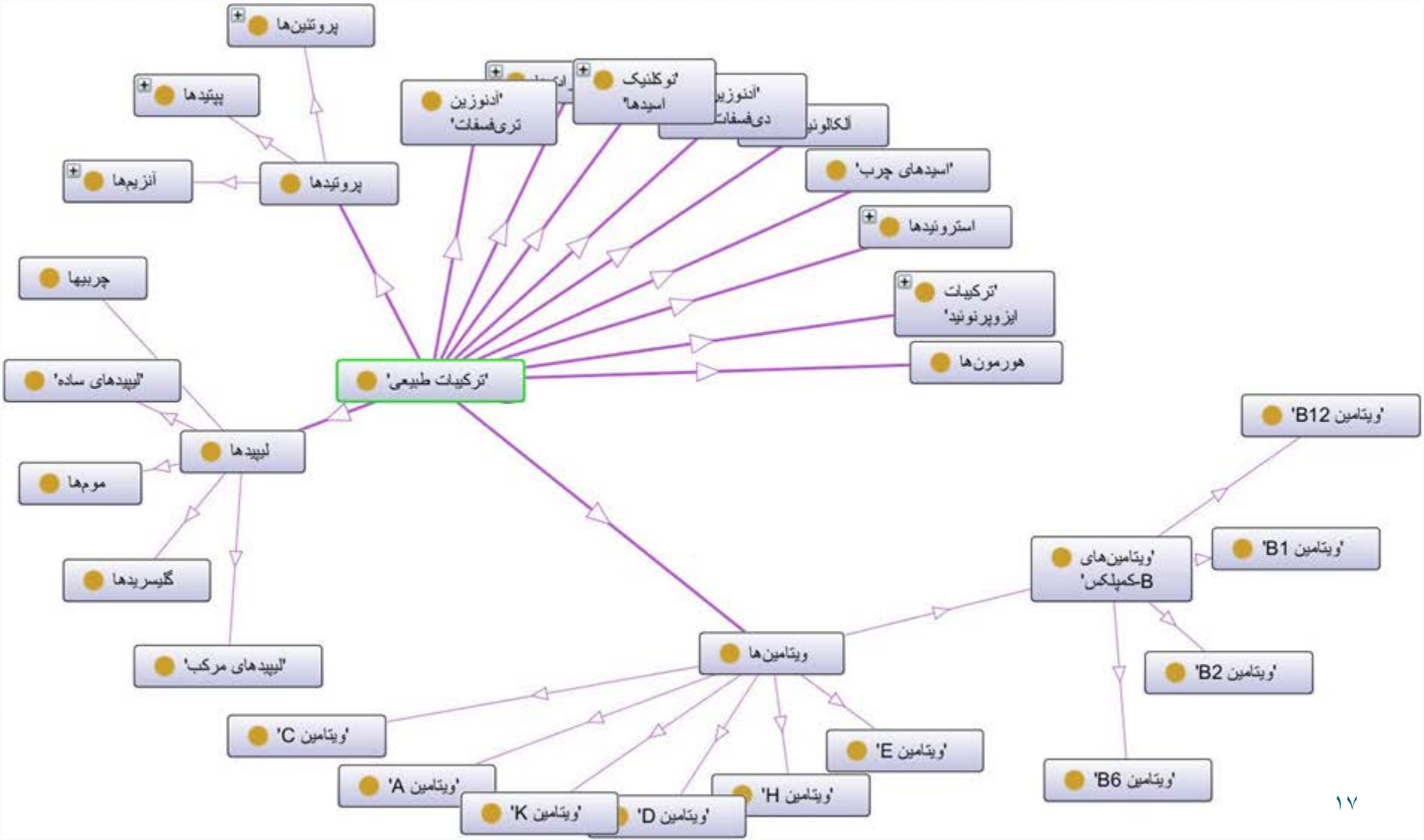
Part of Concept Taxonomy 2



Part of Concept Taxonomy 3



Part of Concept Taxonomy 4



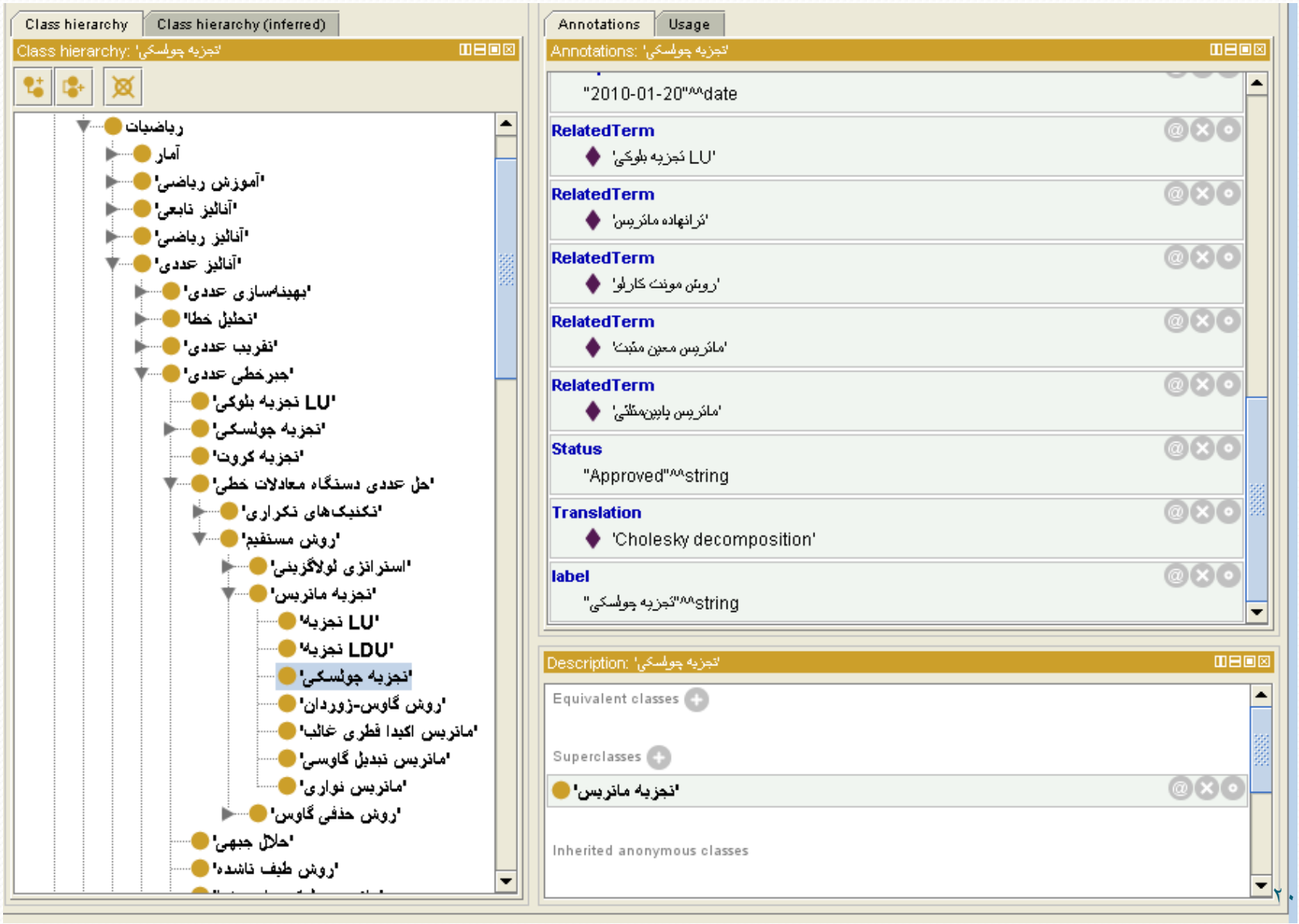
Building Concept Dictionary

- Concept dictionary mainly includes the concept instances for each concept, and their ad hoc relationships.
  - Non-descriptors (non-preferred terms) are identified as **individuals**.
  - Each non-descriptor is identified as an instance of a concept which is designated by its corresponding descriptor.
  - Translations, and abbreviations are set out as properties.

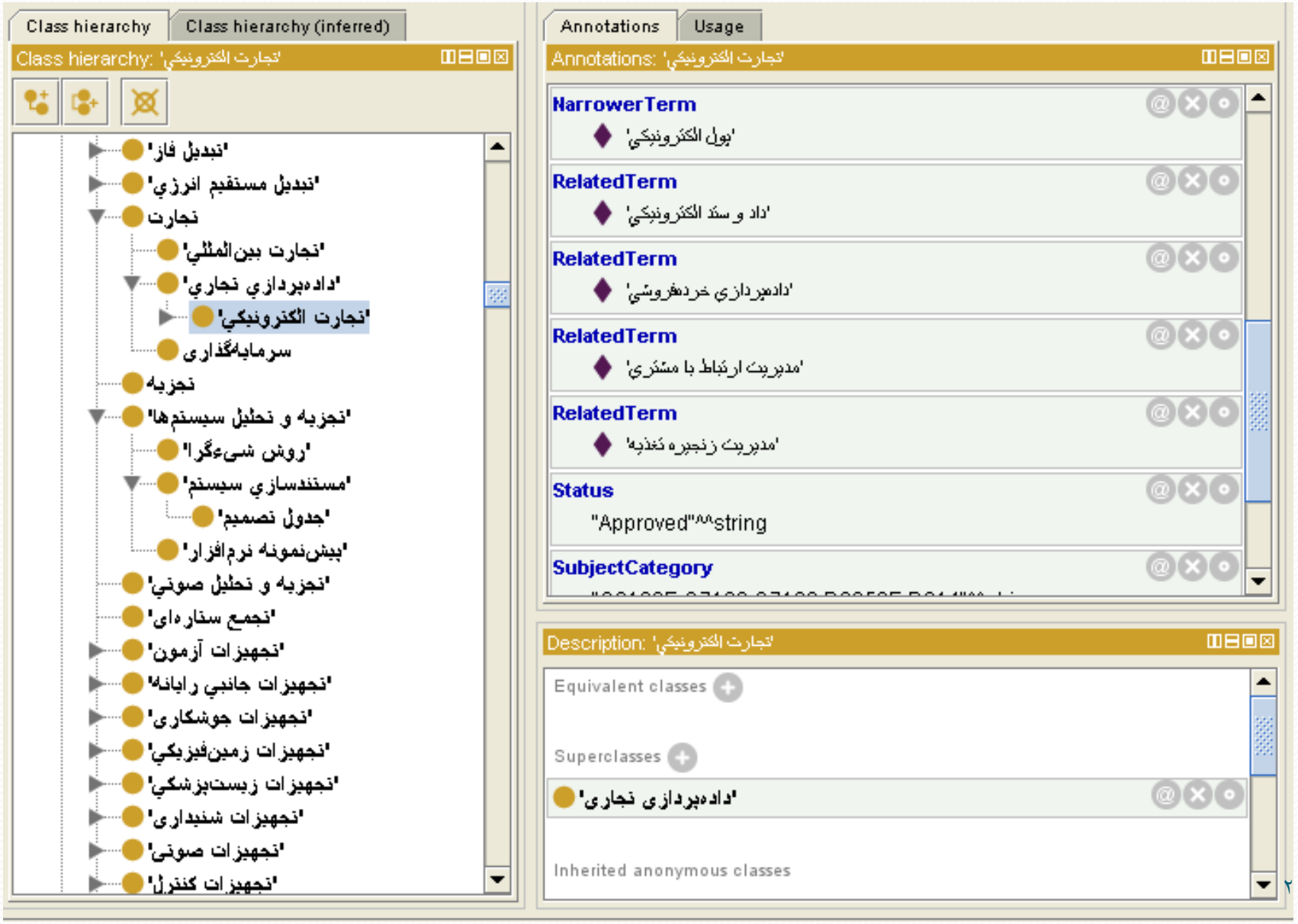
Ontology Implementation

- Preparing Data
  - Contemplating concepts and making common concepts uniform
  - Examining hierarchical and associative relationships and equalizing them semi-automatically
- Implementing conceptual model in OWL (Web Ontology Language)
- Refining ontology by domain experts

Part of Developed Ontology



Another Part of Developed Ontology



Ontology Refinement

- Refinement of semantic relationships is required for extracting ontological relationships from a thesaurus
  - Hierarchical relationships
    - Narrower Term (NT)
    - Broader Term (BT)
  - Associative relationships
    - Related Term (RT)

Refining Relationships in Ontology

Thesaurus Relationships	Possible Ontology Relationships
NT/BT	<hasMember>/<memberOf> <includesSpecific>/<isa> <spatiallyInclude>/<spatiallyIncludeIn> …
RT	<similarTo> <growsIn>/<EnvironmentForGrowing> <treatmentFor>/<treatedWith> <hasMember>/<memberOf> …

"Some relationship constraints", Soergel et al (2004)

Concept Relationships: Examples

- **X, Y are concepts**
- ***Isa***
  - X <*includesSpecific*> Y / Y <*isa*> X
  - X <*inheritsTo*> Y / Y <*inheritsFrom*> X
- ***Whole-part relationships***
  - X <*containsSubstance*> Y / Y <*substanceContainedIn*> X
  - X <*hasIngredient*> Y / Y <*ingredientOf*> X
  - X <*madeFrom*> Y / Y <*usedToMake*> X
  - X <*yieldsPortion*> Y / Y <*portionOf*> X
  - X <*spatiallyIncludes*> Y / Y <*spatiallyIncludedIn*> X
  - X <*hasComponent*> Y / Y <*componentOf*> X
  - X <*includesSubprocess*> Y / Y <*subprocessOf*> X
  - X <*hasMember*> Y / Y <*memberOf*> X

Further Relationship Examples

- **X, Y are concepts**
  - X <*causes*> Y / Y <*causedBy*> X
  - X <*instrumentFor*> Y / Y <*performedByInstrument*> X
  - X <*processFor*> Y / Y <*usesProcess*> X
  - X <*beneficialFor*> Y / Y <*benefitsFrom*> X
  - X <*treatmentFor*> Y / Y <*treatedWith*> X
  - X <*harmfulFor*> Y / Y <*harmedBy*> X
  - X <*hasPhase*> Y / Y <*phaseOf*> X
  - X <*growsIn*> Y / Y <*growthEnvironmentFor*> X
  - X <*hasProperty*> Y / Y <*propertyOf*> X
  - X <*similarTo*> Y / Y <*similarTo* > X
  - X < *oppositeTo* > Y / Y < *oppositeTo* > X

"Concept relationships: Examples", Soergel et al (2004)

Logical generic relationship: Examples

- **X <includeSpecific> Y / Y <*isa*> X**
  - All instances of subclass are also instances of its superclass
  - The subclass inherits all properties of the superclass which is used for hierarchical inheritance

Ontology	Thesaurus
انواع شیمیایی خاک <includeSpecific> خاک های شور chemical soil type <includeSpecific> saline soils	NT
سلول ها <includeSpecific> سلول های خون cells <includeSpecific> blood cells	NT

Part-whole family of relationships: Examples

- **X <containsSubstance> Y / Y <substanceContainedIn> X**
  - Y is the material or substance of which X is made by nature
    - Y loses its identity once it is incorporated into X
  - Whole-part relationship

Ontology	Thesaurus
خون < containsSubstance > لیپیدهای خون Blood < containsSubstance > blood lipids خون < containsSubstance > پروتئین های خون Blood < containsSubstance > blood proteins خون < containsSubstance > سلول های خون Blood < containsSubstance > blood cells	NT

Part-whole family of relationships: Examples

- **X <hasIngredient> Y / Y <ingredientOf> X**
  - Y is the material or substance of which X is made by man
    - Y loses its identity once it is incorporated into X

Ontology	Thesaurus
نوشیدنی های کاکائویی <hasIngredient> پودر کاکائو cocoa beverages <hasIngredient> cocoa powder	**RT**

Part-whole family of relationships: Examples

- **X <hasComponent> Y / Y <componentOf> X**
  - X is a part of Y that retains its identity as an object even when built into the whole
    - each X must be enumerable or nameable

Ontology	Thesaurus
خیش <hasComponent> تیغه های خیش Plough <hasComponent> ploughshare	**RT**
(ساختار سلول) هسته <hasComponent> کروموزوم nucleus <hasComponent> chromosome	**NT**

Part-whole family of relationships: Examples

- **X <spatiallyIncludes> Y / Y <spatiallyIncludedIn> X**
  - X is an inalienable part of Y, identifiable but not inherently separable
  - used for objects with spatial extent

Ontology	Thesaurus
آسیا <spatiallyIncludes> شرق آسیا Asia <spatiallyIncludes> East Asia شرق آسیا <spatiallyIncludes> چین East Asia <spatiallyIncludes> China	NT

Part-whole family of relationships: Examples

- **X <includeSubprocess> Y / Y <subprocessOf> X**
  - processes in thesaurus could be linked using this relation

Ontology	Thesaurus
چرخه آ. تی. پی. <includesSubprocess> فسفریل دار شدن (کردن) ATP cycle <includesSubprocess> phosphorylation	**RT**

Part-whole family of relationships: Examples

- **X <hasMember> Y / Y <memberOf> X**
  - Indicates a relation of membership within a collective or group or organization.

Ontology	Thesaurus
آفریقای فرانسوی زبان <hasMember> بنین Francophone Africa <hasMember> Benin	**RT**
گیاهان آفتکش <hasMember> آرتمیزیا آبسیتیوم Pesticide crops <hasMember> Artemisia absynthium	**RT**

Further relationship examples

- **X <causes> Y / Y <causedBy> X**

Ontology	Thesaurus
فزون چرایی <causes> بیابان زایی overgrazing <causes> desertification	**RT**

- **X <instrumentFor> Y / Y <performedByInstrument> X**

Ontology	Thesaurus
خیش <instrumentFor> شخم زنی plough <instrumentFor> ploughing	**RT**

Further relationship examples

- **X <processFor> Y / Y <usesProcess> X**

Ontology	Thesaurus
تزریق به خاک <processFor> بارورسازی soil injection <processFor> fertilization	**RT**

- **X <hasProperty> Y / Y <propertyOf> X**

Ontology	Thesaurus
گردش خون <hasProperty> فشار خون blood circulation <hasProperty> blood pressure	**NT**

Conclusion

- We used thesaurus as a resource to build ontology
  - Transformation of thesauri from ISO 5964 into ISO 25964
  - Synchronization and Integration of basic sciences thesauri semi-automatically
- Building conceptual model based on semantic information in thesauri
- Generating ontology of basic sciences in OWL
- Refining relationships to more specific ontology relationships by domain expert

References

- Asanee, K. et al (2005), *Automatic term relationship cleaning and refinement for AGROVOC, Workshop on the sixth agricultural ontology services, Portugal.*
- Daconta M. C., et al (2003), *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management,* Wiley Publishing.
- Gómez-Pérez A., et al (2004), *Ontological Engineering*, Springer-Verlag, London.
- Gómez-Pérez A., et al (2010) *A Pattern-Based Method for Re-Engineering Non-Ontological Resources into Ontologies,* International Journal on Semantic Web and Information Systems, 6(4), 27-63.
- Gordon, Fabien L. (2010) *Ontologies in Computer Science: These New "Software components" of Our Information Systems,*INRLA, France

References

- Khosravi, F. and Vazifedoost, A.(2008) *Creating a Persian Ontology through Thesaurus Reengineering for Organizing the Digital Library of the National Library of Iran,* Faslname Ketab,70, 19-36.
- Shamsfard M. et al (2006) *An Ontology Based System for Ranking Documents*, International Journal of Computer Science, Vol 1, No 3.
- Shamsfard, M. et al (2010) *Ontology Population and Semantic Annotation of Persian Texts*, Computer Science and Engineering (CSI)
- Soergel, D. et al (2004) *Reengineering thesauri for new applications: the AGROVOC example*, Journal of Digital Information, Vol. 4, No.4 .
- Sowa, John F. (2010) *The Role of Logic and Ontology in Language and Reasoning, In: Theory and Applications of Ontology . Philosophical Perspectives(chap.11)*, ed. By R. Poli and J. Seibt, Berlin: Springer,2010,231-263.
- Weinbrenner, S. and Engler, J. (2011) *SCY Ontology and metadata scheme*, DIV. 2. 2011, SCY consortium.

Creating Ontology from Persian Thesauri

Molouk Sadat Hosseini Beheshti

Beheshti@irandoc.ac.ir

Faculty Member and Researcher

Iranian Research Institute for Information Science and Technology (IRANDOC)

Fatemeh Ejei

f_ejei@alum.sharif.edu

Sharif University of Technology