Paradigmatic Similarities in Knowledge Representation between KO and AI

Jian Qin

School of Information Studies

Syracuse University

Syracuse, New York, USA

NKOS 2019

Seoul, Korea

Why study paradigms in KR?

Paradigm:

a pattern or model, an exemplar; a typical instance of something, an example.

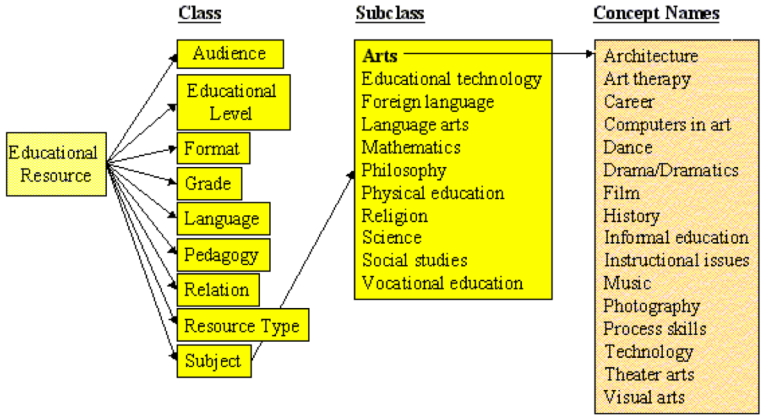
(Oxford English Dictionary)

Simply curious about

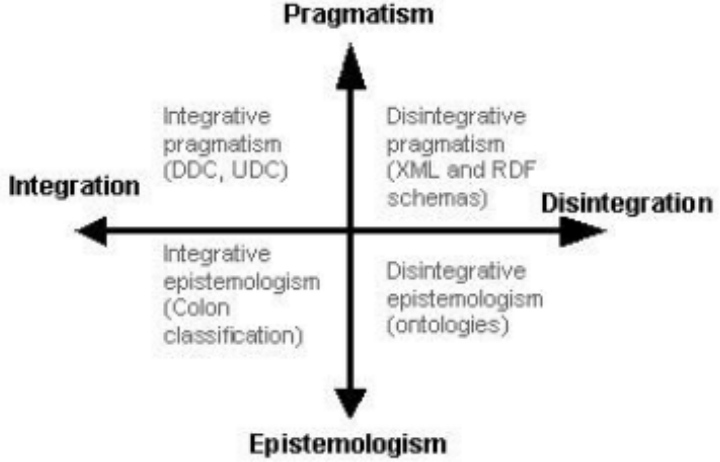
- What is knowledge representation (KR) in AI?
- Can anything be learned from AI's version of KR for us working in KO field?
- KR paradigms in AI context make a great starting point

A personal path to the convergence of KR in KO and AI

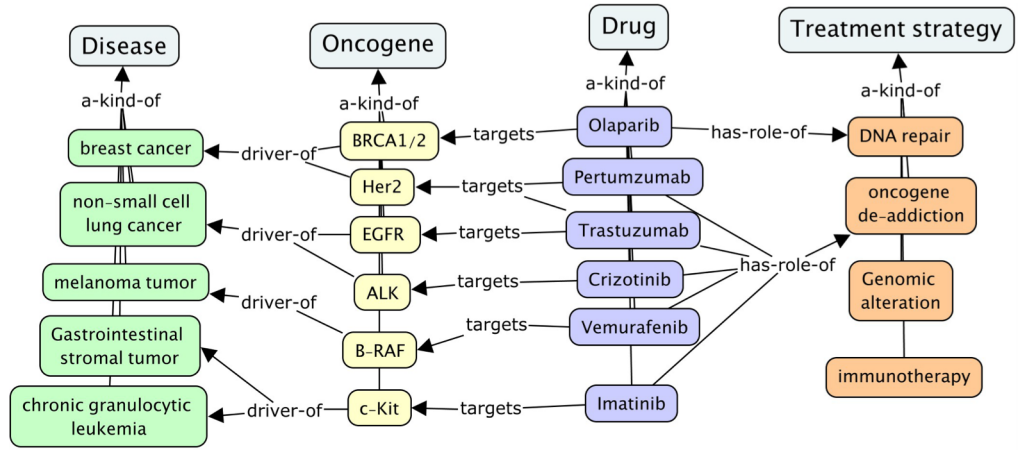
2001: Learning object representation



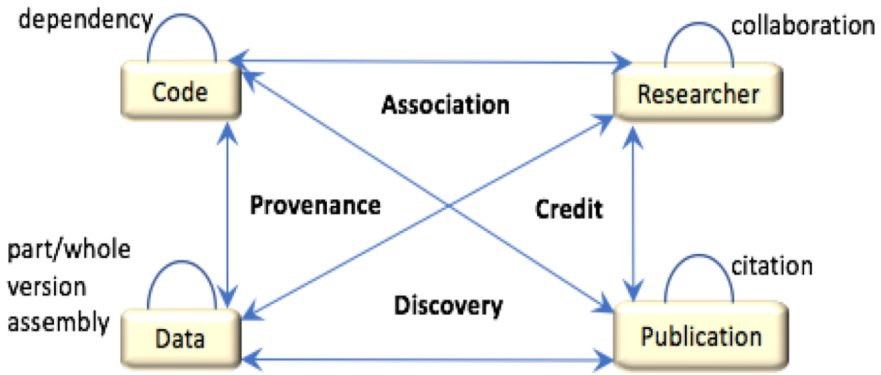
2002: Paradigms in KO



2017: Knowledge node and relation detection from text



2018: A framework for studying relation typology for research data



Disclaimer

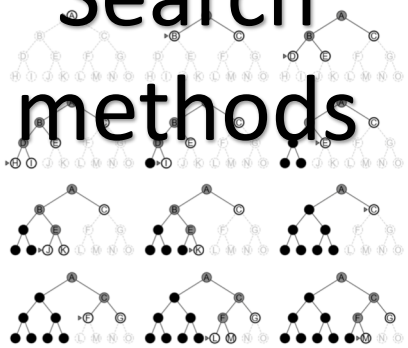
Not a lecture about AI

but instead,

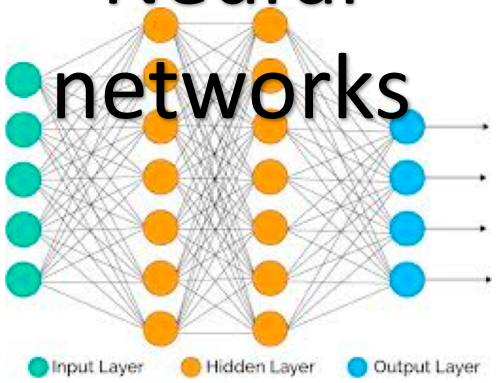
the focus is on paradigmatic similarities in KR between KO and AI

If you heard AI, you may also have heard…

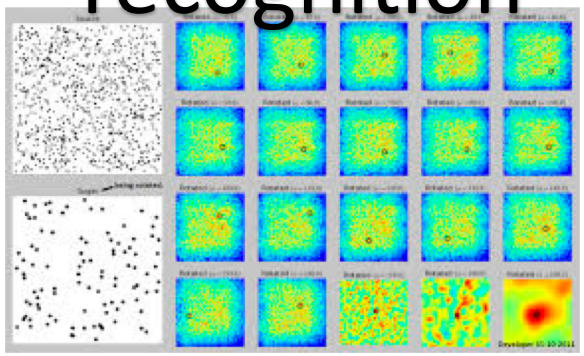
Search methods



Neural networks

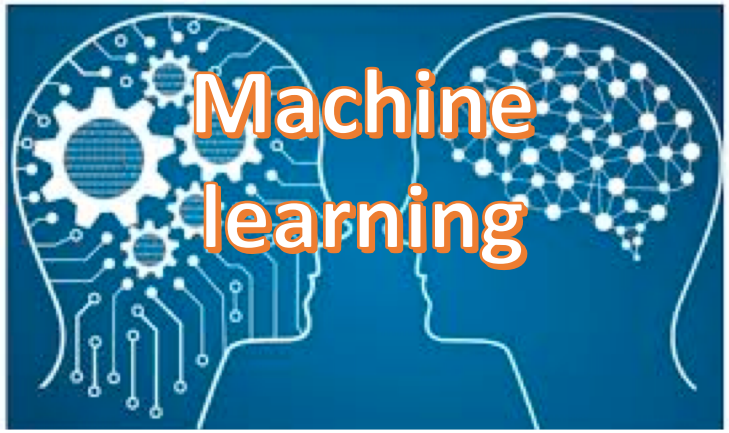


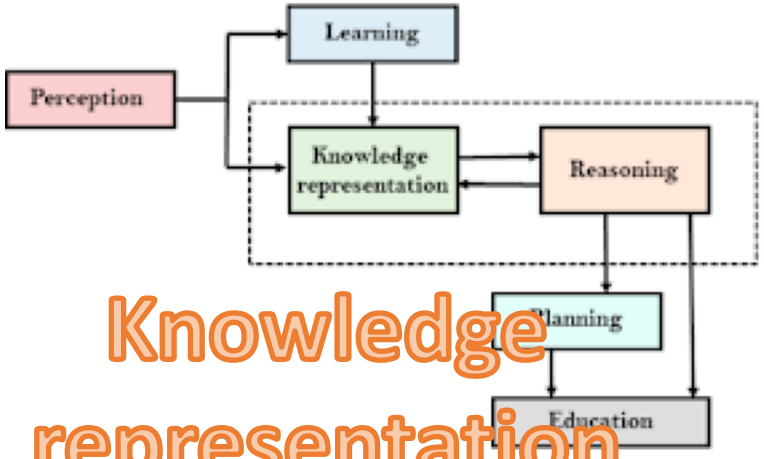
Pattern recognition



First-order predicate logic, propositional logic

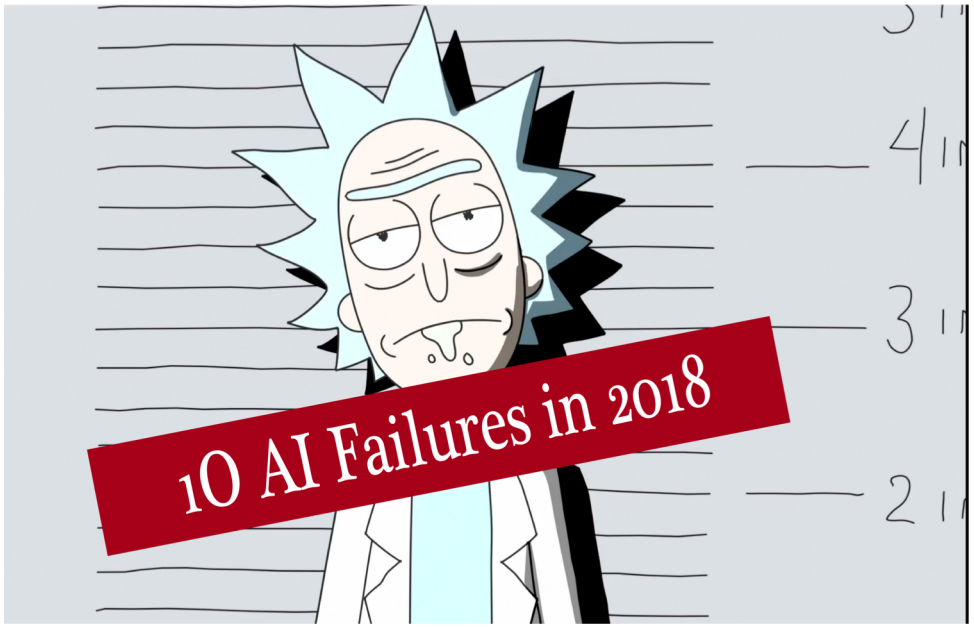
Reasoning with Uncertainty





Knowledge representation

In data-centric, algorithm-dominant AI, there are also bad news



Synced. (2019). 2018 in Review: 10 AI Failures. https://medium.com/syncedreview/2018-in-review-10-ai-failures-c18faadf5983

Uber self-driving car kills a pedestrian

IBM Watson comes up short in healthcare (customers identified "multiple examples of unsafe and incorrect treatment recommendations")

Amazon AI recruiting tool is gender biased

AI World Cup 2018 predictions almost all wrong ("No matter how good your models are, they are only as good as your data…)

Startup claims to predict IQ from faces ("That's phrenology. You just made the ML equivalent of a racist uncle.")

Problems

Epistemological-ontological: concerning the theory-ladenness of data and measurement and how epistemic interests shape big data ontologies.

Epistemological-logical: related to the logical limits of algorithms in modelling clinical complexity and resulting issues of reliability and interpretability for clinical decision making.

Phenomenological: surrounding the irreducibility of human experience to quantitative data that prevents integration into current AI technologies.

Chin-Yee, B. & Upshur, R. (2019). Three Problems with Big Data and Artificial Intelligence in Medicine. Perspectives in Biology and Medicine, 62(2): 237-256.

Inspirations from AI failures and problems

What's at stake in AI for knowledge organization?

What does KO community need or do to ride on the AI wave?

What roles may KO play in sailing through the AI wave?

- Observer? Contributor? Broker? Innovator? Visionary?

How may KO contribute and innovate in the age of AI?

AI research focus areas in the last 30 years

Knowledge engineering

- Model-based reasoning
- Expert systems
- Knowledge-based systems
- Knowledge acquisition
- Knowledge representation
- Knowledge management

Ontologies and terminologies

- Terminological systems
- Ontologies
- Semantic Web
- Semantic technology
- Knowledge representation

Natural language processing

- Text mining

Machine learning, data mining

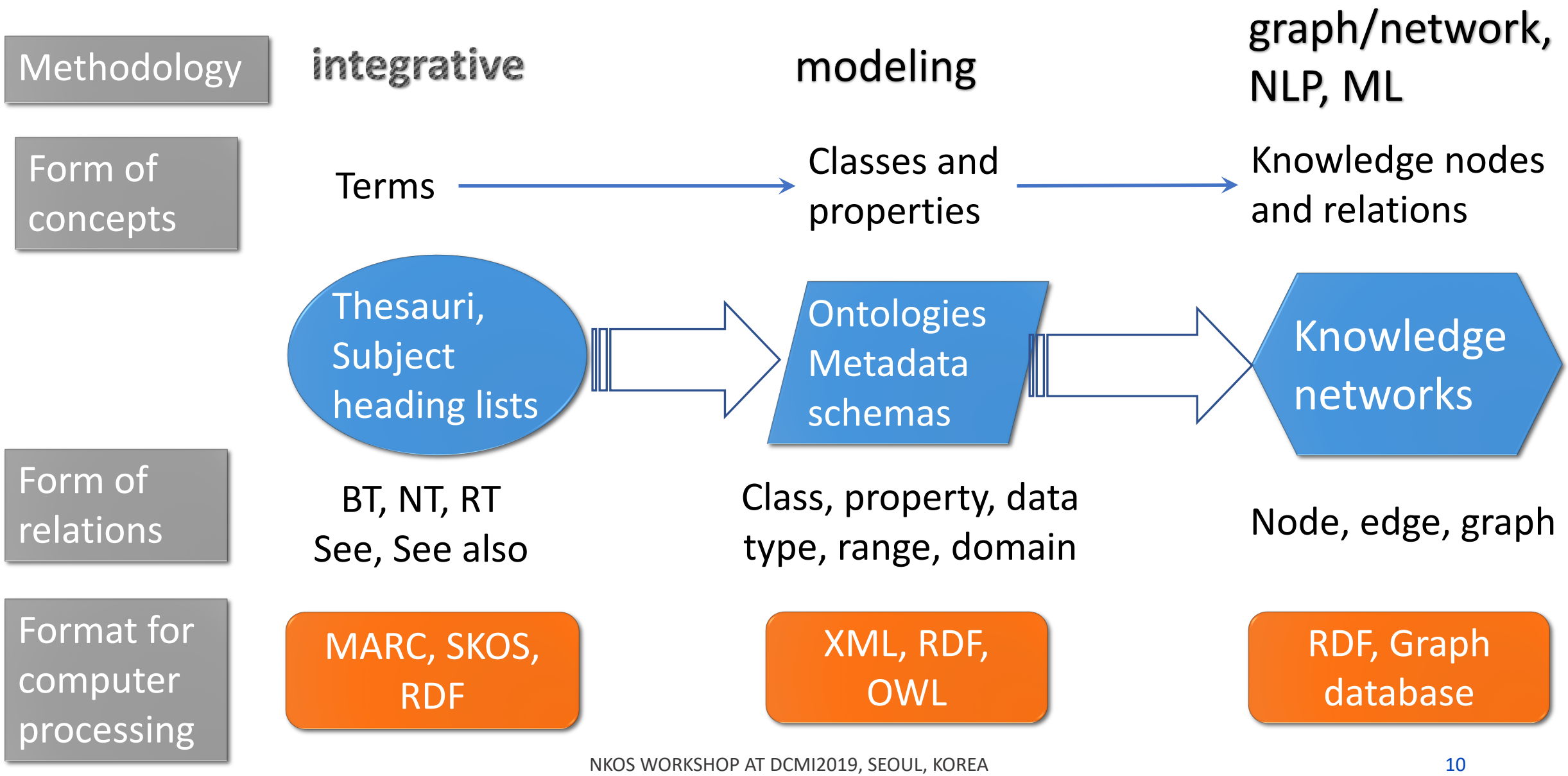
- Pattern recognition
- Classification
- Neural networks
- Clustering
- Data mining

Uncertainty management

- Probabilistic graphical models
- Fuzzy systems
- Decision theory

Peek, N., Combi, C., Marin, R., & Bellazzi, R. (2015). Thirty years of artificial intelligence in medicine (AIME) conferences:A review of research themes. Artificial Intelligence in Medicine, 65: 61-73 .

A developmental view of KOS



A paradigm shift, convergence in representing knowledge

graph/network, NLP, ML

Knowledge nodes and relations

Node, edge, graph

To achieving the vision of knowledge networks:

KO community: What paradigms exist in knowledge organization?

AI community: What paradigms exist in knowledge representation?

Where have the two parallel communities crossed boundaries?

What are some of the converging points between the two communities?

Knowledge organization

KO

as a process targets
to **organize**

Knowledge

in the forms of vocabularies
and classification systems

that are used to
represent
knowledge in

Data
Documents
(and more recently)
Things

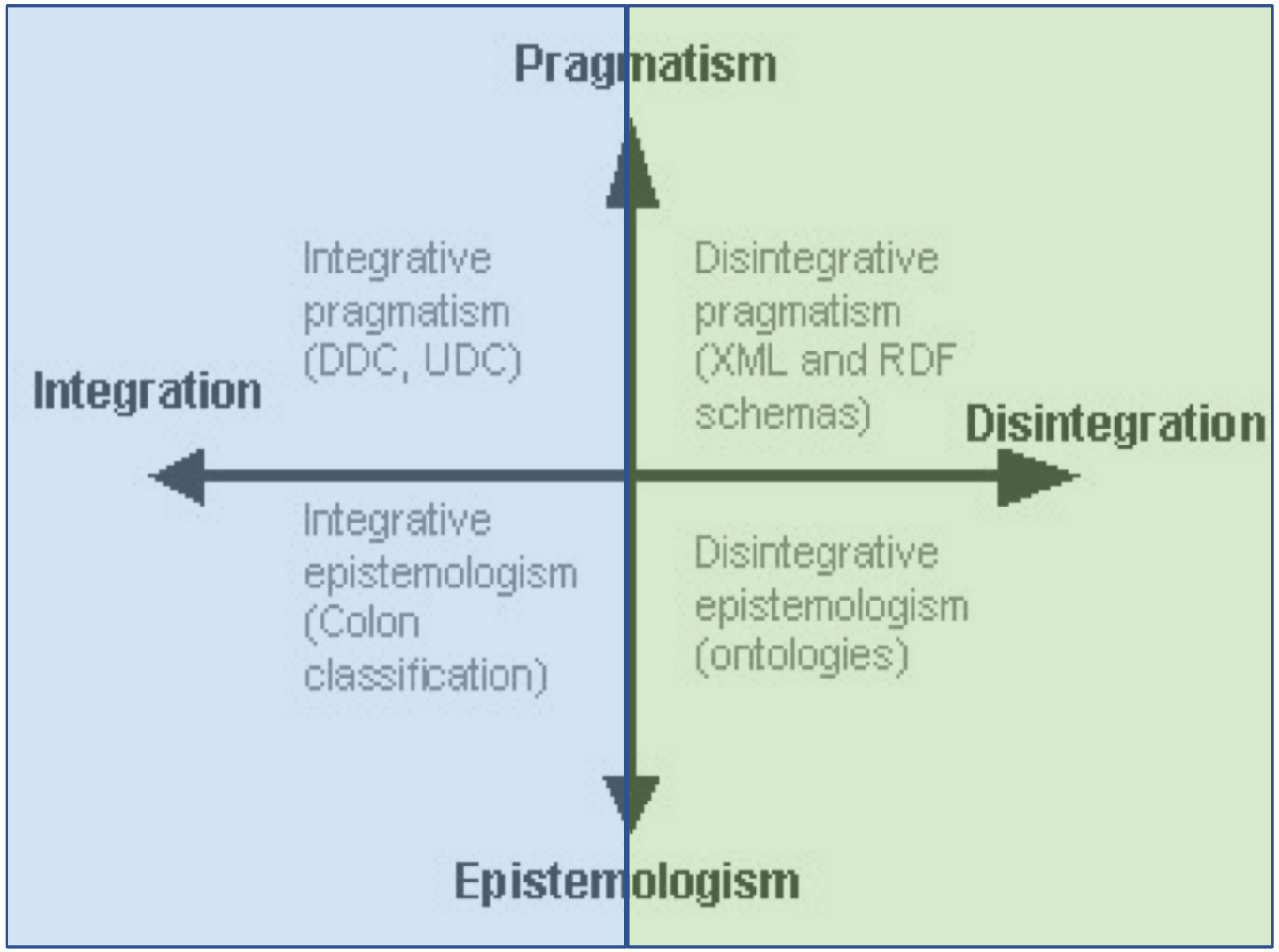
Look through an AI lens

- Not designed to solve specific problems, hence very limited in reasoning
- Only conceptual and/or linguistic relations between terms
- Not intended to represent the fine granular level of human or machine activities

Paradigms in KO

Integration paradigm

Goal is to organize the universe of knowledge into structures or systems, or simply put, KOS is the goal.



Disintegration paradigm

Goal is to model knowledge for a specific domain / application by using KO as a means.

Paradigms in KO

Structures

- Hierarchy / Tree
  - Classification schemes
  - Taxonomy
- Faceted / object-oriented
  - Faceted classification
  - Ontologies
- Alpha-hierarchical
  - Thesauri
  - Subject headings

Relations

- Structured-based
  - Broad term
  - Narrower term
- Semantic-based
  - Related term
  - Variant term
- Inferential-based
  - sameAs
  - Equivalent
  - …

Models and Frameworks

- CIDOC
- BIBFRAME
- …

KOS

Collections of Concepts / Terms in some structures and relations

What is KR in AI?

"a set of ***syntactic*** and ***semantic conventions*** that makes it possible to describe things" --(Bench-Capon 1990, 11)

Representing the real world to allow computer to utilize the knowledge to solve problems and executing tasks, e.g.,

- diagnose a medical condition
- communicating with humans in natural language

Knowledge: how do we talk about it

Mary knows that…

-- can be true /false, right / wrong

Mary fears that…

-- same content, different attitude

Other forms of knowledge

- Know how, who, what, when, …
- Sensorimotor: typing, riding a bicycle, …
- Affective: deep understanding

Belief: not necessarily true and/or held for appropriate reasons

-- "Mary suspect that…"

- Facts about things in our world
- Events or actions that occur in our world
- Knowledge about how to do things
- Knowledge about what we know
- Truths about the real world
- Knowledge rules

Knowledge representation

Symbols standing for things in the world

 → Parking

 → First aid

 → Woman

"John" → John

"John loves Mary" → The proposition that John loves Mary

Knowledge representation:

Symbolic encoding of propositions believed (by some agent)

Principles of KR

- Representation language: *sufficiently precise notation* that it can be *used in, or by, a computer program*;
- Representation scheme: must meet the criteria of *adequacy* ***and*** *expressiveness*
  - Adequate number of representations of physical objects
  - Semantically unambiguous, uniform, notationally convenient, relevant, and declarative

KR paradigms (1)

Production rules (also “symbolic paradigm”): the representation of knowledge as a set of condition action pairs

```
PREMISE          ($AND(SAME CNTXT INFECT PRIMARY-BACTEREMIA)
                       (MEMBF CNTXT SITE STERILESITES)
                       (SAME CNTXT PORTAL GI))
ACTION           (CONCLUDE CNTXT IDENT BACTEROIDES TALLY .7)
If    (1) the infection is primary-bacteremia, and
      (2) the site of the culture is one of the sterilesites, and
      (3) the suspected portal of entry of the organism is the gastro-
          intestinal tract,
then there is suggestive evidence (.7) that the identity of the organism is bacteroides.
```

Davis, R., Buchanan, B., & Shortliffe, E. (1977). Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence*, 8(1): 15-45.

<predicate function> <object> <attribute> <value>

IF *conditions* **THEN** *actions*

condition: tests on WM ``Water is Hot''

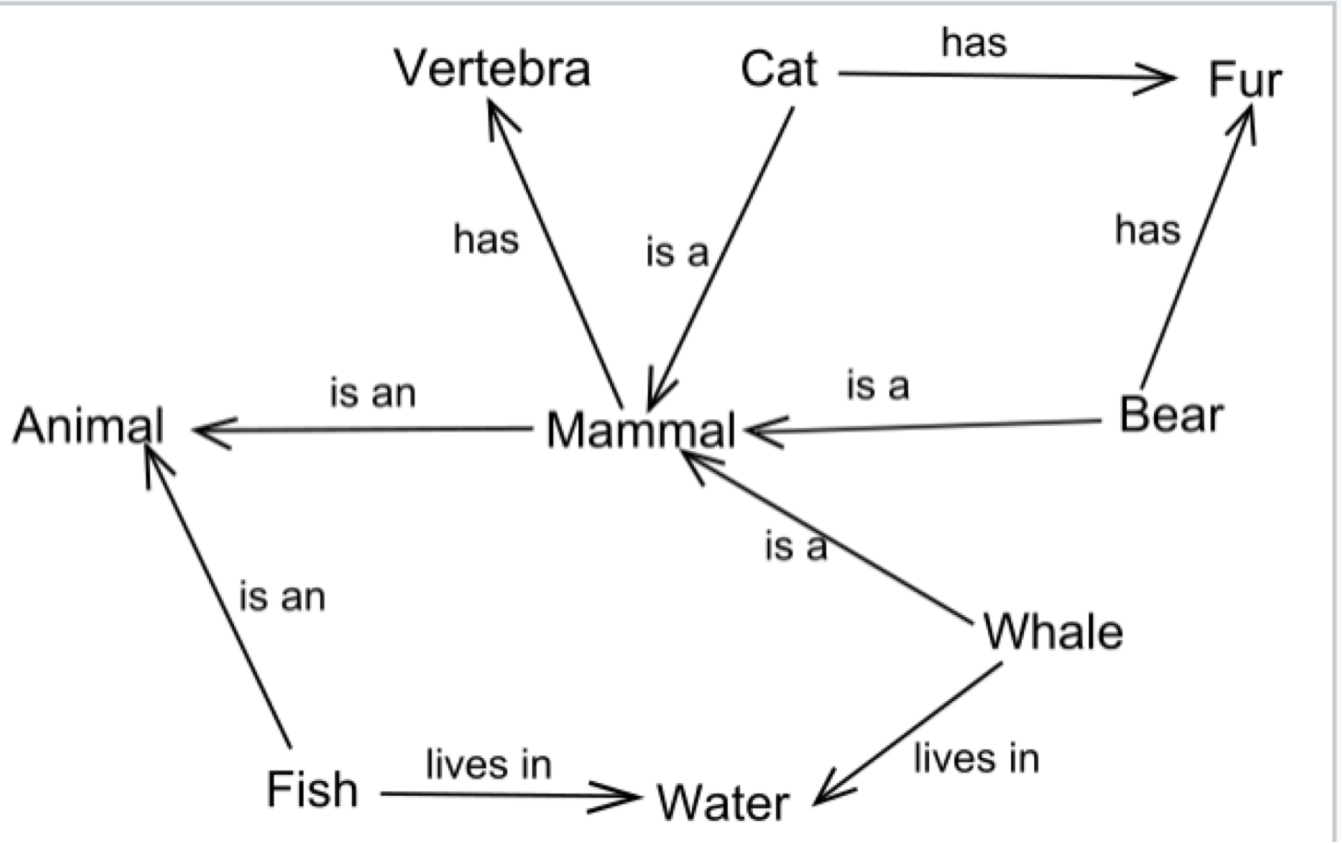
action: changes to WM ``Turnoff the heat''

(Example source: https://web.stanford.edu/class/cs227/Lectures/lec09a.pdf)

KR paradigms (2)

Semantic networks (or simply nets) and frames:

rooted in efforts to build systems to understand natural language by structuring objects in graphs (in mathematics) or networks and object-oriented frames.



(Example source: https://en.wikipedia.org/wiki/Semantic_network

KR paradigms (3)

First-order predicate calculus (also first-order logic): makes statements about the world which are true or false if the state of the affairs it represents is or is not the case.

Premises:

1. If x is a parent of y, then x is older than y
2. If x is the mother of y, then x is a parent of y
3. Lulu is the mother of Fifi

Conclusion:
Lulu is older than Fifi

Mapping to first-order logic:

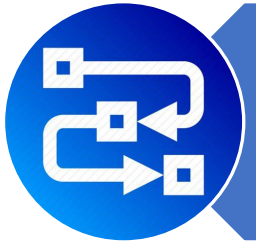
1. $\forall x \forall y. Parent(x, y) \Longrightarrow$ Older(x, y)
2. $\forall x \forall y. Mother(x, y) \Longrightarrow$ Paremt(x, y)
3. Mother (Lulu, fifi)

Conclusion:
Therefore, Older (x, y)

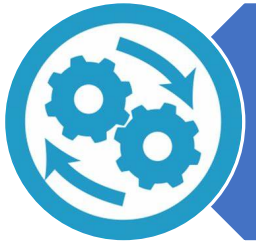
(Example source: https://www.cs.rochester.edu/~brown/173/lectures/logic/formal_logic/FOLogic.html)

Framework for comparison

Goals



Methods



Functions

Similarities in KO and KR paradigms

Paradigm		Goals	Methods	Functions
KO	**Integration**	Organize the knowledge universe	Categorize, classify generalize, synthesize	Represent knowledge in publications and organize knowledge about nature and/or society
	Disintegration	Organize the knowledge in a domain	Categorize, classify, generalize, synthesize, model	Represent knowledge in data and publications in a domain
KR	**Production rules**	Represent knowledge in condition-action pairs to solve problems	Use forward chaining algorithms to execute condition-action pairs	Represent fragmentary knowledge in premise-entity-attribute-value format
	Semantic networks and frames	Represent semantic relations between concepts	Express semantic relations in triples	Connect knowledge nodes through attributes or slots to form a knowledge graph
	First-order logic	Formalize qualifier construction in natural language	Express declarative propositions using the first-order logic syntax and semantics	Produce a set of axioms for reasoning

An experiment: knowledge node and relation detection

What: Detecting knowledge nodes and relations in biomedical publications

Why: Gain insights into the types of knowledge nodes and relations

How: Manual coding and automatic generation of knowledge nodes and relations

```
LOCUS       SCU49845     5028 bp    DNA
DEFINITION  Saccharomyces cerevisiae TCP1-beta gen
            (AXL2) and Rev7p (REV7) genes, complet
ACCESSION   U49845
VERSION     U49845.1  GI:1293613
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeas
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharo
            Saccharomycetales; Saccharomycetaceae;
```

Example: Taxonomic representation of a DNA sequence dataset in GenBank that documents the organism in the form of taxon lineage

Example: Knowledge representation of the publication related to the DNA sequence dataset (PubMed ID: 7871890), which strives to provide as many topical access points as possible

MeSH Terms

Amino Acid Sequence
Base Sequence
Chromosomes, Fungal
Cloning, Molecular
DNA Damage*
DNA Replication
DNA, Fungal/biosynthesis
DNA, Fungal/secretion
DNA-Directed DNA Polymerase*
Fungal Proteins/chemistry
Fungal Proteins/genetics*
Genes, Fungal*
Genetic Complementation Test
Molecular Sequence Data
Mutagenesis*
Open Reading Frames
Saccharomyces cerevisiae/chemistry
Saccharomyces cerevisiae/genetics*
Saccharomyces cerevisiae Proteins*
Sequence Analysis, DNA
Sequence Homology, Amino Acid

Substances

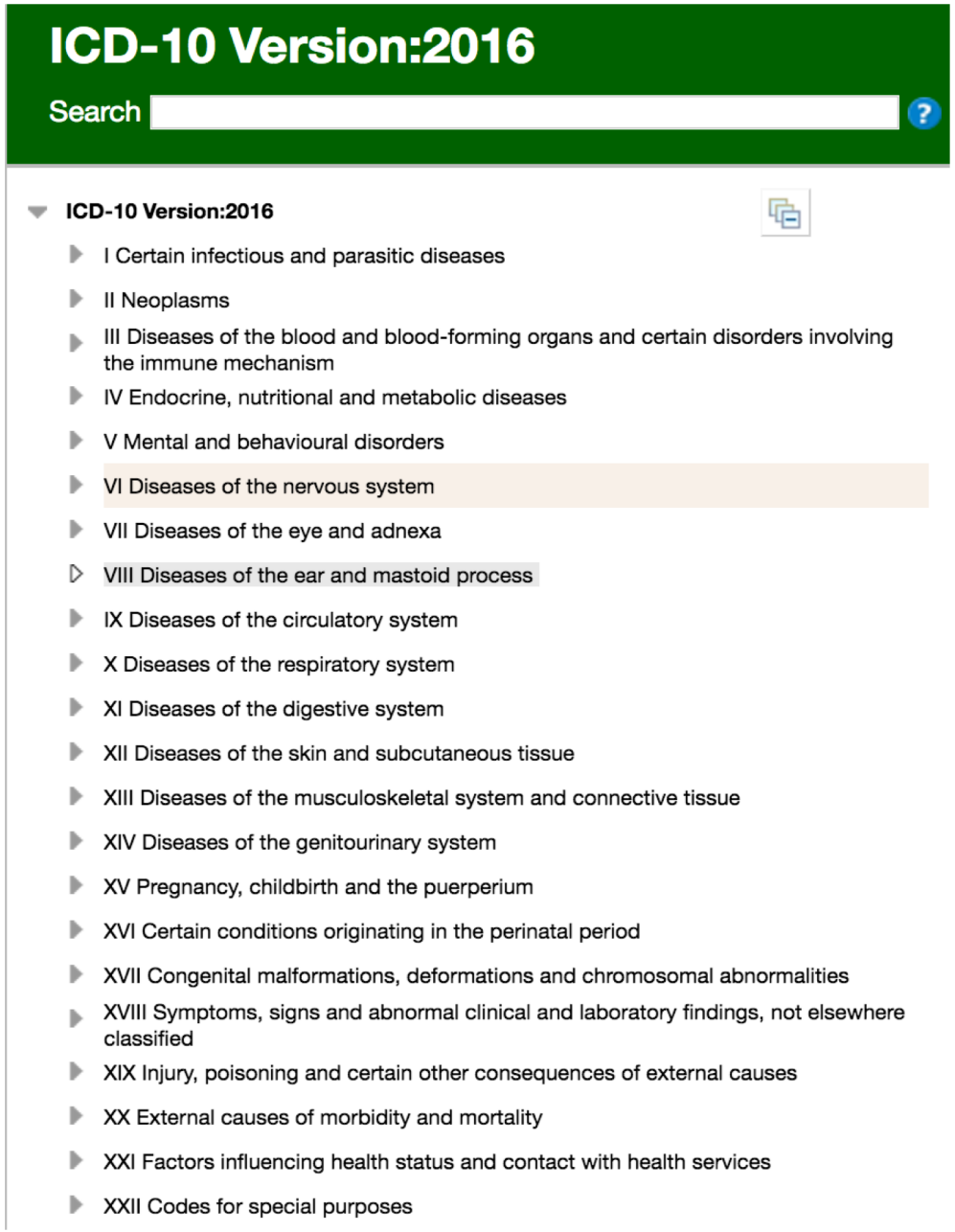
DNA, Fungal
Fungal Proteins
REV7 protein, S cerevisiae
Saccharomyces cerevisiae Proteins
DNA-Directed DNA Polymerase

Representation gaps between “bench science” and “bedside practice”

Traditional representation of knowledge is focused on concepts, i.e., “points” or “terms”

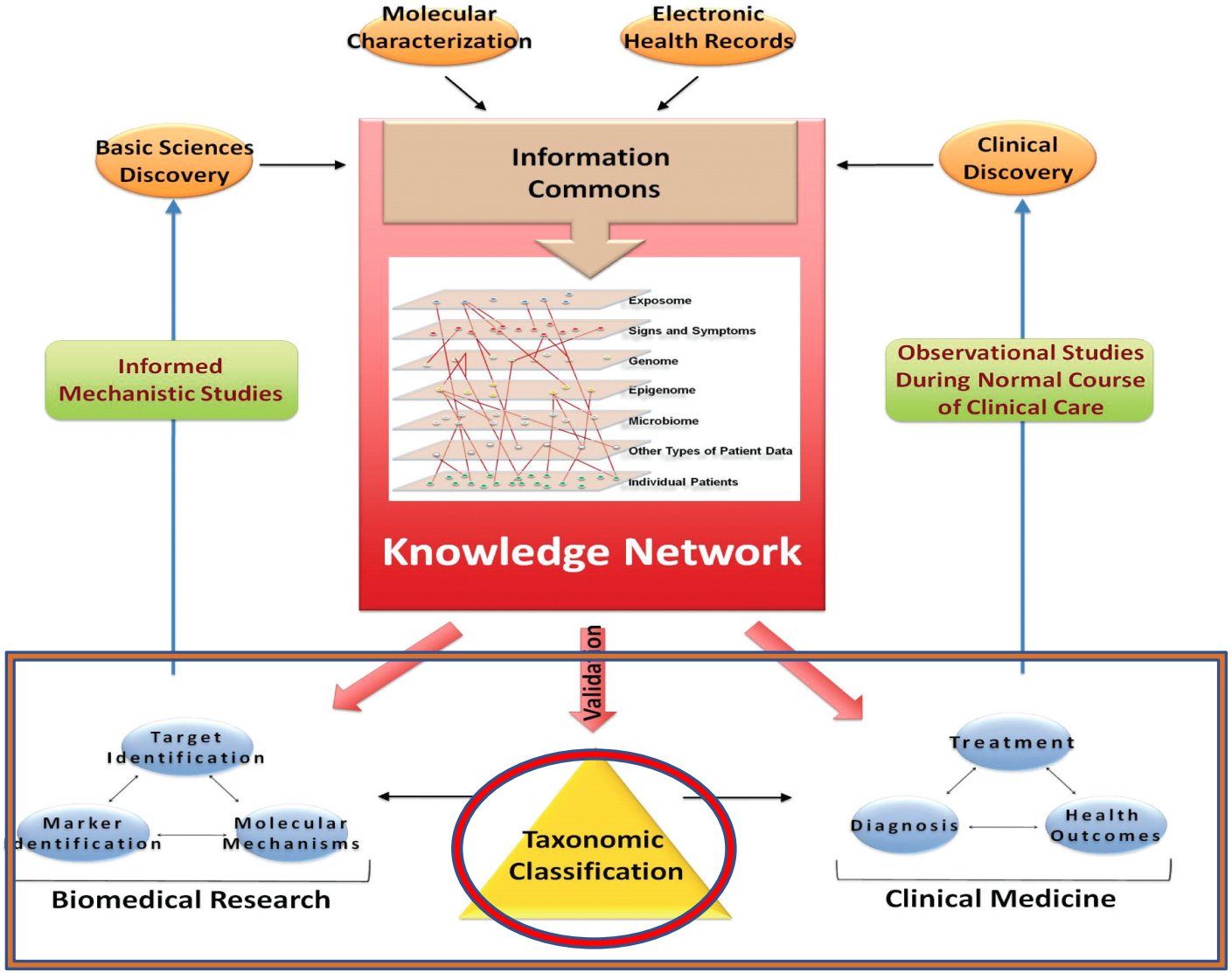
- Disease-centric
- Often coarse, less refined
- Only simple relationships are present

Language differences in clinical and basic research lab settings



Source:
http://apps.who.int/classifications/icd10/browse/2016/en

The vision of a Knowledge Network of Disease and Information Commons



(Source: NAS, 2011)

Linking data to research publications as a way to bridge knowledge representation gaps

New information and discovery is reported in

Attributes of data (metadata)

Object-to-object linking
Concept-to-concept linking
Label-to-term linking
Node-to-node linking

Detect from publications

Knowledge nodes:
- Types?
- Attributes?

Relationships between nodes:
- Types?
- Attributes?

Pilot Study : Data and Methods

A sample of 30 articles in precision medicine

- Four in Breast Cancer
- Five in Diabetes
- Eleven in Oncology

- "Purposeful Sampling"

- To gain insights and in-depth understanding rather than empirical generalizations

Pilot study: what to detect

- *Molecular entities* such as genes, proteins, genomes, etc.
- *Disease names*
- Names or terms related to *treatments/therapies*
- Methods, techniques, and types of decisions related to *diagnosis*
- *Data sources* used by the publication
- *Types of relations* between potential knowledge nodes

Structural levels of nodes

Examples of knowledge nodes manually derived from the sample publications

Category	Atomic Level (name of things)	Concept Level	Cluster Level
Gene	HER2, BRCA1, BRCA2, EGFR	Oncogenes	EGFR mutations in lung cancer
Disease	Non-squamous carcinoma, squamous cell carcinoma	Non-small cell lung cancer	Lung Cancer
Drug	Pertumzumab, Lmatinib, Crizotinib	Tyrosine kinase inhibitor	Oncogene de-addition

Pattern nodes and relations (1)

Knowledge nodes may be categorized by

- Disciplinary field:

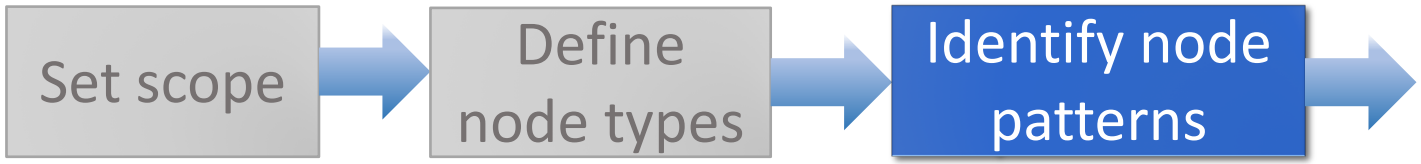
  genetics, pathology, pathophysiology, oncology, virology, …

- Disease name and biomarker pairs:
  - Chronic myeloid leukemia (CML) with mutated gene BCR-ABL
  - Breast cancer with positive estrogen receptor (ER), BRCA1/2, and Her2
  - Non-small cell lung cancer with mutations in multiple genes such as epidermal growth factor receptor (EGFR), excision repair-cross complementation group (ERCC), and ribonucleotide reductase (RRM)

Pattern nodes and relations (2)

Knowledge nodes that blend **clinical** and **basic research**

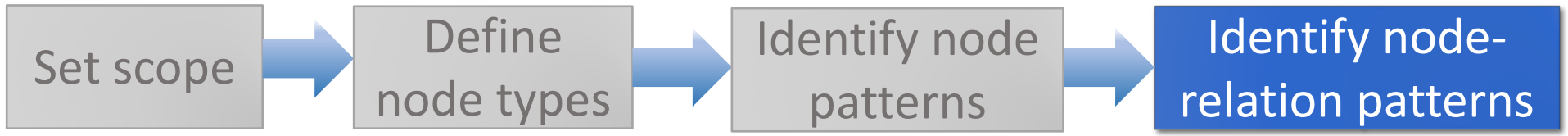
- clinically actionable mutations
- phenotype of breast cancer
- resistance to endocrine therapy
- biomarkers predicting response to therapy
- genomic drivers of cancer
- predictive and prognostic biomarkers
- intratumor heterogeneity
- molecular classification of tumors



K-node-relation patterns

Major relation types and patterns between knowledge nodes observed in the sample publications

Relation	Pattern	Example
has-biomarker	Disease has–biomarker Gene	chronic myeloid leukemia *has biomarker* BCR-ABL non-small cell lung cancer *has-biomarker* EGFR
Is-driver-of	Gene is-driver-of Disease	HER2 *is-driver-of* breast cancer C-Kit *is-driver-of* chronic granulocytic leukemia
targets	Drug targets Gene	Crizotinib *targets* ALK Olaparib *targets* BRCA1/S
has-role-of	Drug has-role-of Treatment	Crizotinib *has-role-of* oncogene de-addition Olaparib *has-role-of* DNA repair



What was learned from this project

- Knowledge nodes may be marked with different labels—structure, discipline, disease, gene or biomarker, treatment, …
- Each label represents a dimension and the nodes in one dimension form a vector
- A node may reside in multiple dimensions at the same time
- Building a knowledge network in any domain needs
  - Knowledge acquisition: NLP + ML + KOS + human intelligence
  - Knowledge representation: which paradigm? Much is unknown

Looking forward

Potential for KO to contribute to KR in AI:

- Assisting knowledge acquisition in KR processes, e.g.,
  - Remodeling KOS to develop knowledge bases
  - Using KOS to build relation vocabularies and rules

KR thinking in KO:

- Digital data and information calls for disruptive innovations in KO practices
- Inter-influence between KO and KR: extending the power of human intelligence to enrich and enhance artificial intelligence

Thanks and

Questions?

The Quad view of Syracuse University