

„Where is the wisdom we have lost  
in knowledge?

Where is the knowledge we have  
lost in information?

T. S. Eliot (1888-1965)



# Knowledge Discovery Methods An Exploration of Different Approaches

Stefan Smolnik  
University of Paderborn



2. Konferenz  
Professionelles Wissensmanagement  
Erfahrungen und Visionen

Luzern, Schweiz, 02. - 04. April 2003

# Overview

1. Context and motivation
2. Knowledge and knowledge discovery
3. Exploration of knowledge discovery methods
  - ⇒ Piaget: the discovery of small children
  - ⇒ Churchman: inquiring systems
  - ⇒ Knowledge discovery in databases and data mining
  - ⇒ Deriving general world knowledge from texts and taxonomies
  - ⇒ Topic maps: knowledge structures and structured semantic link networks
4. Findings
5. Summary and conclusions

## Context and motivation

- Complex and dynamic environments: companies have to create and to renew the resource knowledge continuously
- Necessary prerequisite: availability of knowledge
- Decisive role: creation and discovery of knowledge
- Filling of “knowledge gaps”
- Competitive advantage from a lead of knowledge and the capability to transform superior knowledge into market-driven business processes and decisions.

# Knowledge and knowledge discovery

- Literature: many definitions of knowledge
- Data, information, knowledge, sense, wisdom
- Knowledge: information and person-specific aspects like experiences, values, and insights
- Polanyi: “We can know more than we can tell.”
- Knowledge discovery methods enable an increase of transparency of knowledge in organizations and support users finding relevant knowledge.



# Piaget: the discovery of small children

- How small children learn to discover the world: cognitive development, intellectual development, and discovery processes
- Epistemological explanation of how individuals learn and how knowledge is constructed
- Knowledge: invention, reinvention, construction, reconstruction, reorganization
- Interaction with the world and real world objects; active involvement

# Piaget: the discovery of small children

- ➔ Four stages of intellectual development
  - ➔ Sensori-motor (birth-2 years): cognitive orientation; agents of action; action with intentions; object permanence
  - ➔ Pre-operational (2-7 years): language; representation and classification of objects; egocentric thinking
  - ➔ Concrete-operational (7-11 years): logical thinking; multi-dimensional classifications; simple operations
  - ➔ Formal operations (11 years and up): abstract propositions and systematical tests of hypotheses; propositional, content-free and logic thinking

## Churchman: inquiring systems

- “Cognition creates knowledge.”
  - Inquiring systems: organizations, which are supported by IT since creation and discovery of knowledge are essential for their efficiency in competitive environments
- Purpose: creation of knowledge
  - Philosophical and pragmatic framework: critical processes of discovery of new knowledge and renewal of existing knowledge for better design of organizational KMS



## Churchman: inquiring systems

	Leibnitz	Locke	Kant	Hegel	Singer
Knowledge discovery method	Formal logic  Sentence generator	Assign labels to inputs  Communication	Construct models from theories  Interpret data  Choose best model	Construct theses, antithesis  Dialectic	Strategy of agreement  Sweeping-in
Knowledge	Fact nets  Tautologies  Contingent truths	Taxonomy	Fact Nets	Synthesis	New standard  Exoteric knowledge  Simplistic optimism
IT support	Expert systems	Databases  Networking	Internet  Knowledge bases  Model base	Dialectic engines	Expert systems  Internet/WWW

# Knowledge discovery in databases and data mining

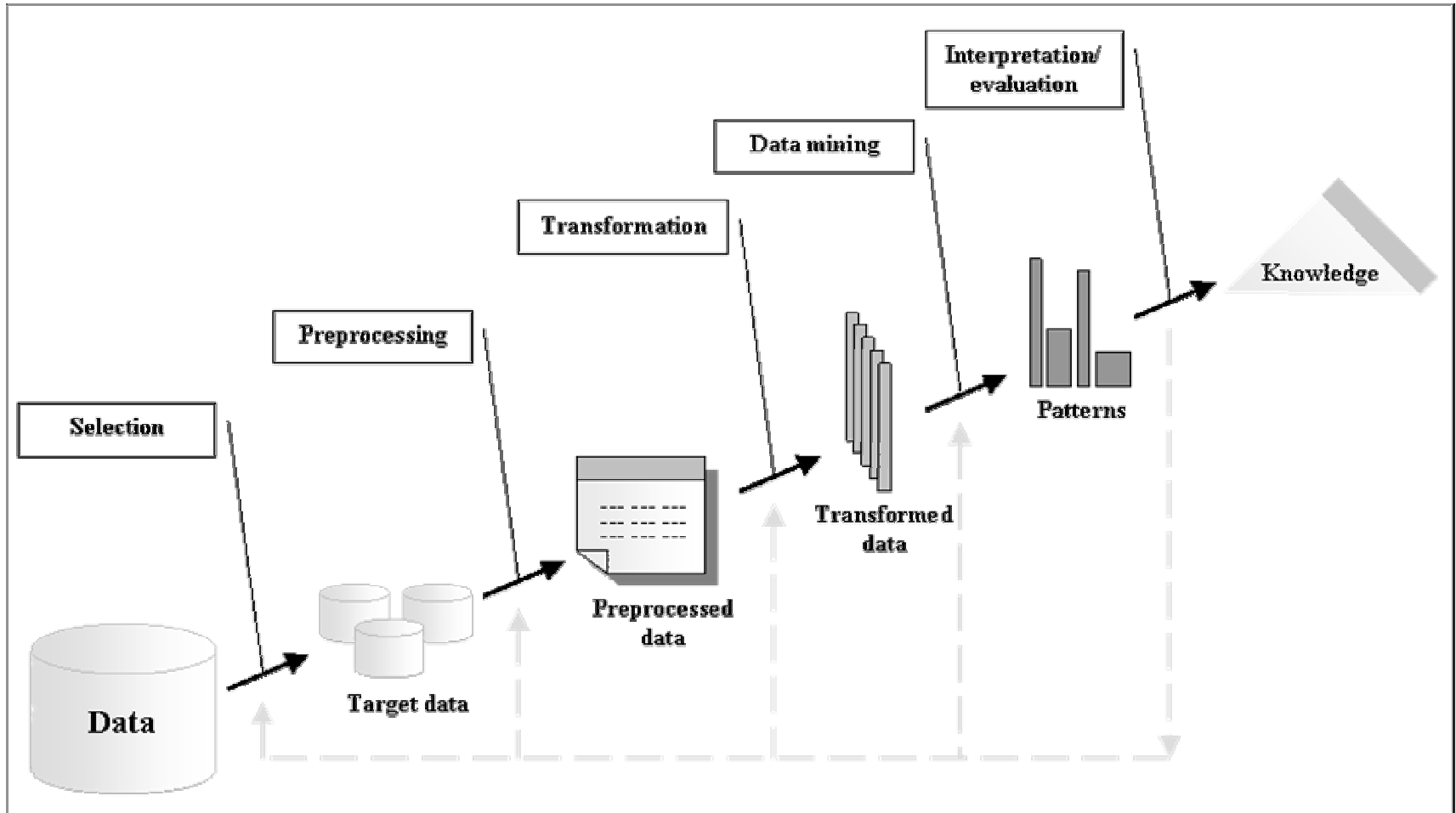
→ “Knowledge discovery in databases (KDD) is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.”

→ The identification of patterns in large structured data sets results in knowledge.

→ The core process of knowledge discovery is data mining.

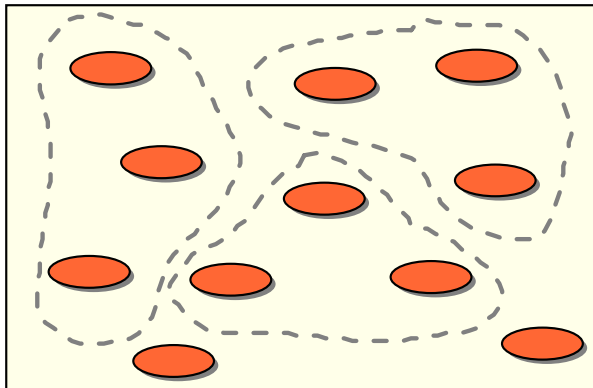
→ Knowledge can be interpreted as discovered relationships between data sets, which have a novel meaning for an individual.

# Knowledge discovery in databases and data mining

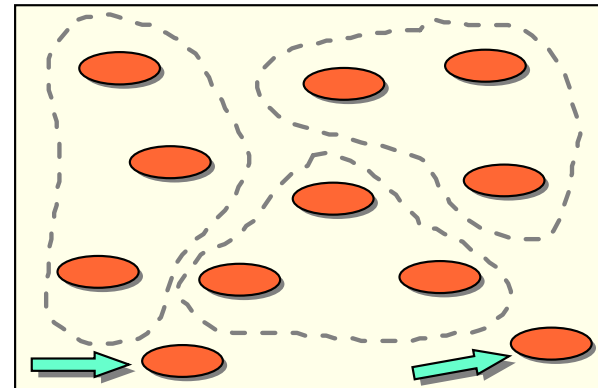


# Knowledge discovery in databases and data mining

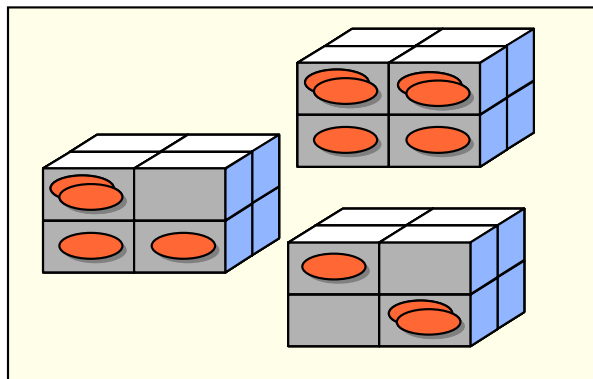
**Clustering**



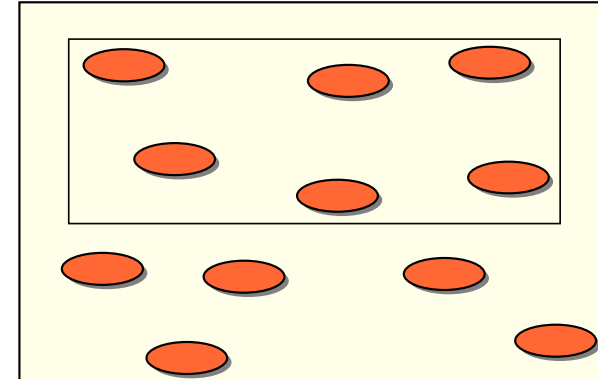
**Deviation detection**



**Classification**



**Summarization**



## Deriving general world knowledge from texts and taxonomies

- ⇒ Schubert: “... that there is a largely untapped source of general knowledge in texts, lying at a level beneath the explicit assertional content.”
- ⇒ Knowledge: experiences, values, and relationships, which are normal, commonplace, or well known in the world
- ⇒ First step: abstraction by simplifying and generalizing
- ⇒ Second step: stronger generalizations

# Deriving general world knowledge from texts and taxonomies

- Taxonomies: structures of categories or topics to which documents, thus texts, can be assigned
  - Ontologies: data constructs that reflect the structure of a body of knowledge by including categories, vocabulary, and information about relationships
- An ontology is generally assembled by humans and then used to analyze documents
  - Common and shared understanding

# Topic maps: knowledge structures and structured semantic link networks

„GPS of the information universe“:

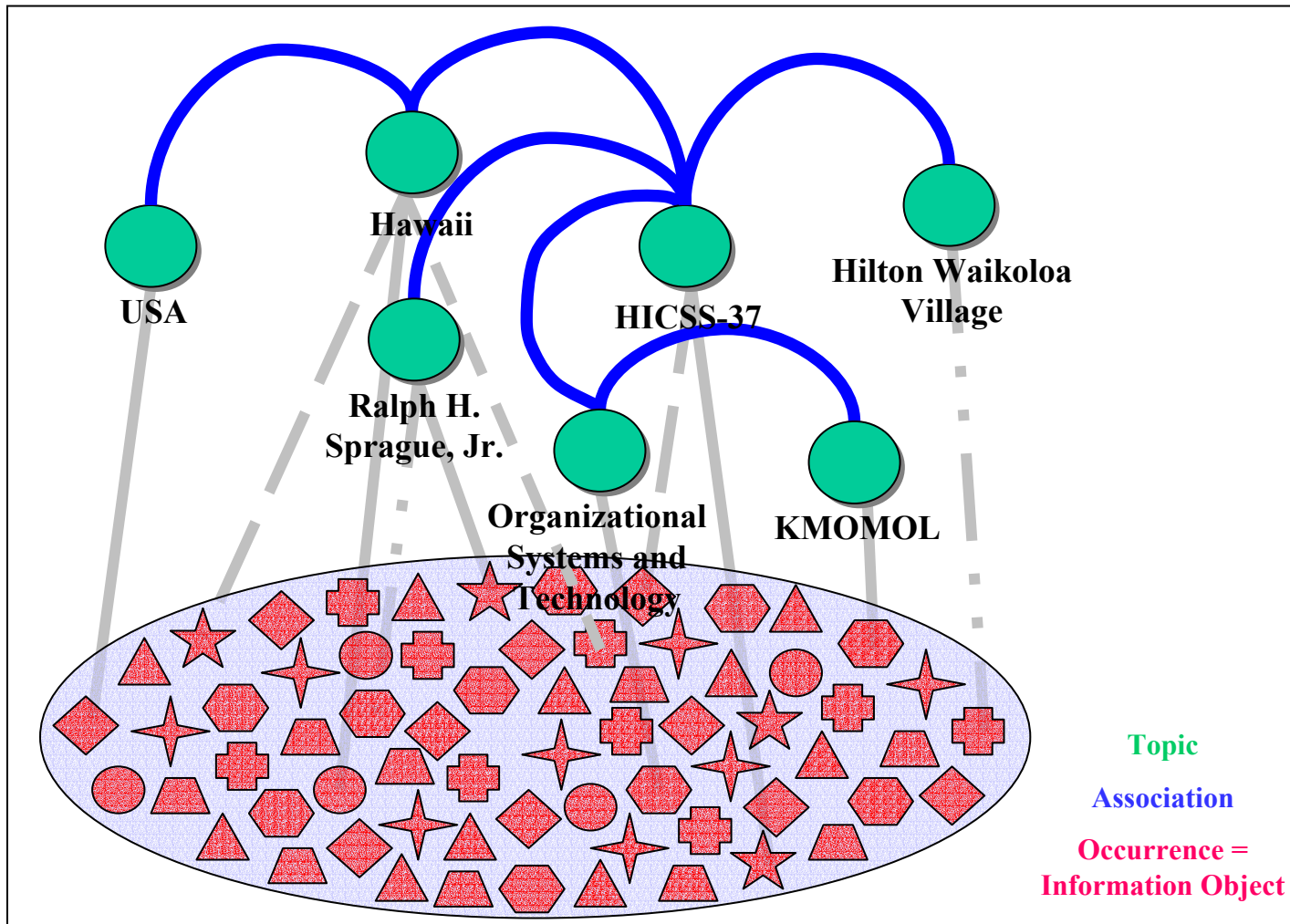
- Applying topic maps to information resources generates knowledge structures
- Structured semantic link network above great sets of information resources
- Searching in a topic map can be compared to searching in knowledge structures
- Base technology for knowledge representation and knowledge management

# Topic maps: knowledge structures and structured semantic link networks

- ➔ The structured semantic link network makes an easy and selective navigation to requested information possible (independent of a real information object)
- ➔ Transparent access to real information objects by searching in knowledge structures, i.e. in a meta layer above the real information objects
- ➔ Added value, i.e. new knowledge through new contexts
- ➔ Representation of ontologies
- ➔ “Bridging the gap” between knowledge representation and the field of information management



# Topic maps: knowledge structures and structured semantic link networks



# Findings

- Activities on objects to create and discover knowledge
- Objects:
  - real world objects
  - standards and procedures
  - organizational memories
  - large sets of structured data, documents and texts
  - large sets of heterogeneous information resources

# Findings

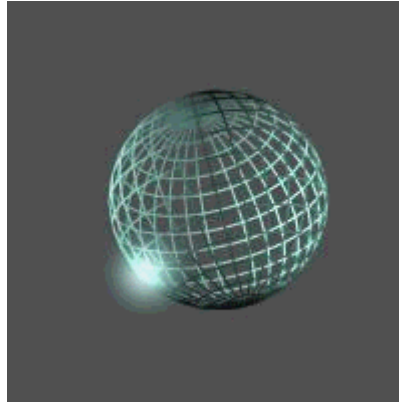
- Activities:
  - cognitive processes like gaining experiences and insights
  - the construction of intrinsic values, personal qualities, and intangible assets
  - the association of objects and the construction of semantic networks of relations

	Discovery of children	Inquiring systems	KDD and data mining	Text analysis and taxonomies	Topic maps
Methods, algorithms, and activities	Observation Cognitive orientation “Learning by doing” Classification Internalization Operation Thinking	Formal logic Communication Classification Construction Interpretation	Clustering Deviation detection Classification Dependency modeling Summarization	Derivation Abstraction Simplification Generalization Ontologies	Associating Structuring Modeling Ontologies
Resulting Knowledge types	Experiences Real world objects and events Acting and thinking processes Causes and effects	Fact nets Tautologies Taxonomies Synthesis New standards Exoteric knowledge	Valid, novel, useful, and Understandable patterns Unsuspected relationships	Experiences, values, and Relationships Normal, commonplace, and well known knowledge Taxonomies	Reusable Structured semantic link networks Semantic relations
Fields of application	Epistemological foundation	Philosophical foundation	Large sets of structured data	Documents and texts	Large sets of Heterogeneous Information resources

## Summary and conclusions

- Comprehensive literature study
- Individual selection of methods
- Intrinsic focus on methods
- Strong conceptual foundation and clarification of meanings
- Similar activities and several types of knowledge

# Questions, remarks, and discussion



K-Discovery @Web: <http://gtm.upb.de>

Stefan Smolnik  
University of Paderborn  
Business Computing 2  
Faculty of Business Administration,  
Business Computing & Economics  
Prof. Dr. Ludwig Nastansky  
Email: [Stefan.Smolnik@notes.upb.de](mailto:Stefan.Smolnik@notes.upb.de)  
GCC @Web: <http://gcc.upb.de>