Artif Intell Rev
DOI 10.1007/s10462-011-9221-2

Modeling semantic information in engineering applications: a review

Kunmei Wen · Yong Zeng · Ruixuan Li · Jianqiang Lin

Abstract Due to the latest advances of information technology and the increasing complexity of engineering applications, it is becoming more and more important to model semantic information. There are many modeling methodologies to do the work of modeling semantic information instead of natural language processing. Since this field is very broad, the comparison discussed here is not an exhaustive study but rather the partial views of the coauthors from our own perspectives. In the present paper we give a review of the literature of conceptual models especially static one and then classify them into four type models namely structure-based model, object-oriented model, knowledge semantic-based model, and web semantic-based model. Based on the classification given above, a hierarchy structured criteria is given. According to the criteria we pick one or two representative conceptual models from each type to conduct the comparison. We compare the following five aspects of conceptual models: expressivity, clarity, semantics, formal foundation, and application fields. The comparative study shows that different models have different features and fit different fields of engineering applications. The present comparison study is useful for users to understand and choose right conceptual models combining with specific requirements of engineering applications.

Keywords Semantic information · Conceptual model · Modeling semantic information

K. Wen (✉) · R. Li
College of Computer Science and Technology, Huazhong University of Science and Technology,
1037 Luoyu Road, 430074 Wuhan, China
e-mail: kmwen@hust.edu.cn

R. Li
e-mail: rxli@hust.edu.cn

Y. Zeng · J. Lin
Concordia Institute for Information Systems Engineering, Concordia University,
1455 de Maisonneuve West, EV.07.633, Montreal, QC, H3G 1M8, Canada

Y. Zeng
e-mail: zeng@ciise.concordia.ca

Published online: 04 May 2011

1 Introduction

Due to the latest advances of information technology and the increasing complexity of design process, there is huge amount of unstructured information. That results in more and more semantic information hidden in engineering applications. Modeling semantic information is very important to effectively and efficiently manage and reuse the information generated during the whole process of engineering applications. Modeling semantic information can also benefit to the inference of engineering knowledge. We can get some extensional information by using reasoning on the semantic information. Definitely it benefits to the communication among multi disciplines or multi organizations. Therefore, modeling semantic information becomes a stable basis for subsequent development of engineering applications.

During the whole process of engineering applications, a large amount of semantic information (Li et al. 2008) is generated. Some of them is captured in form of documents such as reports, notebooks, requirement specifications (Gnesi 2005; Mala 2006) memos, emails, sketches, patent documents (Kanda et al. 2008), product survey, 2D/3D computer-aided design (CAD) elements, and product lifecycle management. At the same time other semantic information is still retained in the memory of the engineers. Most semantic information of engineering applications is unstructured, in contrast to structured data resources such as database tables.

How to model semantic information of these documents? Modeling the semantic information means mapping the documents of engineering applications into conceptual models. It is shown in Fig. 1. Conceptual models serve as a communication tool between computers and users. In this paper, we will make a comparative study of conceptual models, mainly modeling the semantic information of engineering application. The main difference between our work and the existed work is that we conduct the comparison based on our own classification and criteria, by using the idea "equivalent class partitions" in software testing. First, conceptual models are classified from the functional viewpoint. Then based on the requirements for the semantic information, criteria of conceptual models with hierarchy structure are proposed.

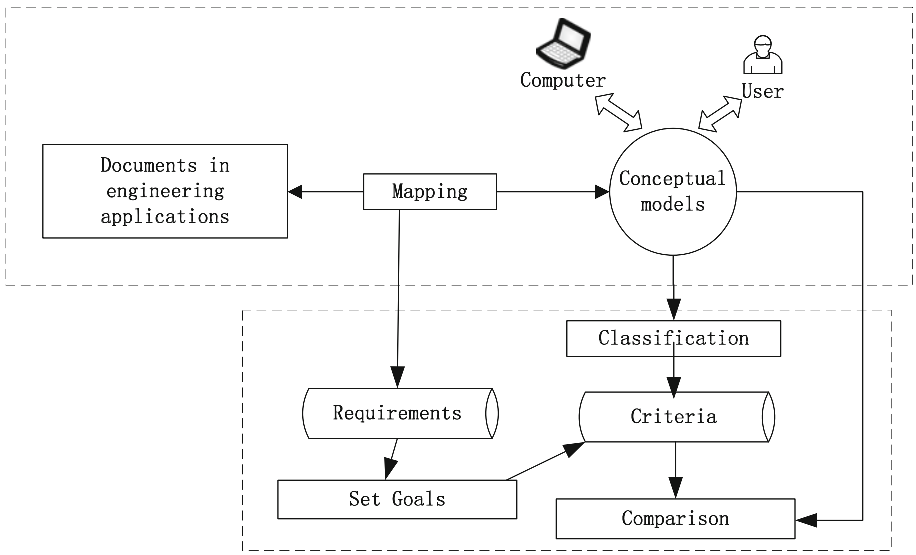

Fig. 1 Structure of the review of modeling semantic information

Finally, we pick one or two representative conceptual models as "equivalent class" from each type. According to the criteria, the comparison is conducted among these representative conceptual models instead of comparing all conceptual models. Therefore, the upper section of Fig. 1 is the foundation of this paper. The lower section of Fig. 1 is the structure of the review of modeling semantic information.

Modeling semantic information means to construct conceptual models in engineering applications. Conceptual models represent concepts and relationships between them. The conceptual model is a well known technique of data model, together with logical model and physical model. The aim of conceptual model is to express the meaning of terms and concepts used by domain experts to discuss the problem, to find the correct relationships between different concepts, and to communicate, abstract and compute. The conceptual model is also intended to be used to specify systems so that stakeholders (e.g., customers, operators, analysts, designers) can better understand the system being modeled. Many forms of symbolic notation have been developed to enable conceptual models to represent various levels of abstraction. Lexical and graphic are the main representation forms of conceptual models.

Conceptual models can be used to specify systems requirements, structures and behaviors. The conceptual model attempts to clarify the meaning of various usually ambiguous terms, and ensure that problems with different interpretations of the terms and concepts cannot occur. It also defines information requirements for an entire industry and provides the basis for industry-wide standardization and development of generic software solutions (Allworth 1999; Fettke and Loos 2003). Conceptual modeling naturally belongs to a sub-discipline of requirement engineering (as conceptual models are used to define user requirements) and software engineering (as conceptual models are used to develop, acquire or modify information systems). Conceptual model plays very important roles over the development lifecycle.

Various kinds of conceptual models are applied in different disciplines, including computer science, information management, business process modeling, software engineering and system engineering. A conceptual model can be described using various notations, such as universal modeling language (UML) and object modeling technology (OMT) for object modeling, IDEF1X for entity relationship modeling. There are many existing conceptual models, each of which has its own advantages and weakness. Nevertheless, to the best of our knowledge, there is little comparison work on different conceptual models at present. This may confuse people how to select a right conceptual model. Hence, how to choose and use these models is very important. A good comparison should play an important role as one of standards of computer industry. The comparison is useful to find out the characteristics and differences of these conceptual models. It is also helpful to figure out which conceptual models are suitable for some applications and which elements are important for choosing models. Therefore based on the comparison users can select the right model to meet their application requirements.

Before the comparison it is necessary to discuss the classification and comparison criteria of conceptual models.

2 Conceptual models and classification

During the past 30 years, many conceptual modeling methods have been proposed for different objectives. It is impossible to compare all the conceptual models. Before the comparison, we will give a review of conceptual models and classify them from the functional viewpoint. The classification from the functional view proposed here is different from the existing classification. Different people may give totally different classifications. For example, from

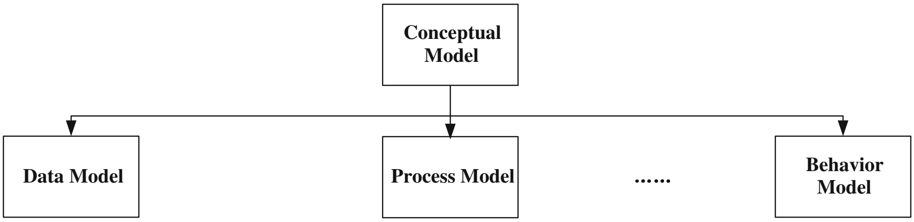

Fig. 2 Types of conceptual models

the view of expression form, conceptual models can be classified into graphic, textual and combinational conceptual models. Conceptual models can also be classified into static and dynamic conceptual models based on different modeling objects. The classification from the functional view is useful to help user select conceptual models based on the functional understanding.

Conceptual models are classified into several types (Becker et al. 1995): data model (e.g. Entity Relationship and Extending Entity Relationship (ER and EER)), process model (e.g. Data Flow Diagram(DFD), petri net and process algebras), behavior model (e.g. state chart diagram) and so on, as shown in Fig. 2. In addition, there are other conceptual models, such as dimensional model used in data warehousing and reference model used as a reference for various purposes.

Data model is static conceptual model while process model and behavior model are dynamic models. We focus on static conceptual model. In this paper, conceptual model represents conceptual data model for the sake of simplicity. Many forms of conceptual data models have been developed to represent various levels of abstraction. We will give the introduction of the following representative models.

ER: One of the earliest, Chen's Entity Relationship Model (Chen 1976), offers a set of shapes and lines which, much like musical notation, deliver a wealth of information with sparse economy of drawing. Chen (1983, 1997) studied the correspondence between English sentences and ER diagrams, and processed eleven rules for translation. Hartmann (2007) revised the several correspondences between English sentence structures and concepts of ER modeling. However given the lack of clarity in definitions, it is not surprising that (Codd 1990) says "The major problem with the entity-relationship approach is that one person's entity is another person's relationship." Date (2006)) also agrees Codd's oppions, saying that the ER approach is flawed because the very same object can quite legitimately be regarded as an entity by some users and a relationship by others.

IE and IEM: Information Engineering (IE) or Information Engineering Methodology (IEM) is an approach for designing and developing information systems. It is said to have originated in Australia between 1976 and 1980, and appears first in the literature in 1981 in the Savant Institute publication 'Information Engineering' by Martin (1991). Information Engineering first provided data analysis and database design techniques.

Petri net: A Petri net (also known as a place/transition net or P/T net) is one of several mathematical representations of discrete distributed systems. As a modeling language, it graphically depicts the structure of a distributed system as a directed bipartite graph with annotations. As such, a Petri net has place nodes, transition nodes, and directed arcs connecting places with transitions. Petri nets were invented in 1962 by Petri in his PhD thesis (Petri 1962). Peterson made Petri nets more spread and gave several detail examples (Peterson

1981). Reisig applied Petri Net theory into concurrent and distributed algorithms (Reisig 1985, 1992).

ORM: Object Role Modeling (ORM) is a powerful method for designing and querying database models at the conceptual level, where the application is described in terms easily understood by non-technical users. In practice, ORM data models often capture more business rules, and are easier to validate and evolve than data models in other approaches. ORM evolved from the Natural language Information Analysis Method, a methodology that was initially developed in Europe in the mid-1970s, and later in the 1980s. Halpin (1995) provided the first formalization of Object-Role Modeling in joint papers and the work, Conceptual Schema and Relational Database Design.

UML: The Unified Modeling Language (UML) is a standard language for specifying (OMG 2007, 2008), visualizing, constructing, and documenting the artifacts of software systems, as well as for business modeling and other non-software systems. The development of UML began in late 1994 when Booch and Rumbaugh of Rational Software Corporation began their work on unifying the Booch and OMT (Object Modeling Technique) methods. In the Fall of 1995, Jacobson and his Objectory company joined Rational and this unification effort, merging in the OOSE (Object-Oriented Software Engineering) method (Rumbaugh et al. 2004; Booch et al. 2005). For data modeling purposes, UML includes class diagrams that may be annotated with expressions in a textual constraint language.

Concept map: The technique of Concept map was developed by Novak and Canas (2008) as a means of representing the emerging science knowledge of students. Concept maps are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts. Words on the line, referred to as linking words or linking phrases, specify the relationship between the two concepts. An important characteristic of concept maps is that the concepts are represented in hierarchical fashion.

Mind map (or similar concepts): A Mind map is a diagram used to represent words, ideas, tasks or other terms linked to an arranged around a central key word or idea. Mind maps (Buzan and Buzan 1996) have been used for centuries, for learning, brainstorming, memory, visual thinking, and problem solving by educators, engineers, psychologists and people in general. Buzan used image-centered radial graphic organization techniques referred to variably as mental or generic mind maps for centuries in areas such as engineering, psychology, and education. Cognitive map, mental map, mind map, cognitive model, or mental modelis is a type of mental processing (cognition) composed of a series of psychological transformations by which an individual can acquire, code, store, recall, and decode information about the relative locations and attributes of phenomena in their everyday or metaphorical spatial environment. Tolman is generally credited with the introduction of the term 'cognitive map' (Tolman 2000). Here, 'cognition' can be used to refer to the mental models, or belief systems, that people use to perceive, contextualize, simplify, and make sense of otherwise complex problems.

XML: The Extensible Markup Language (XML) is a general-purpose specification for creating custom markup languages (W3C 2006). Deitel introduces the XML markup for a book in Harvey and Deitel (2000).

RDF: The Resource Description Framework (RDF) (W3C 2004; Smith and Deborah 2004; W3C 2004) is a language for representing information about resources in the World Wide Web. It is particularly intended for representing metadata about Web resources. RDF Schema (RDFS) (Brickley and Guha 2002) was built by the World Wide Web Consortium (W3C) as an extension to RDF with frame-based primitives. The combination of both RDF and RDF Schema is normally known as RDF(S). Three more languages have been developed

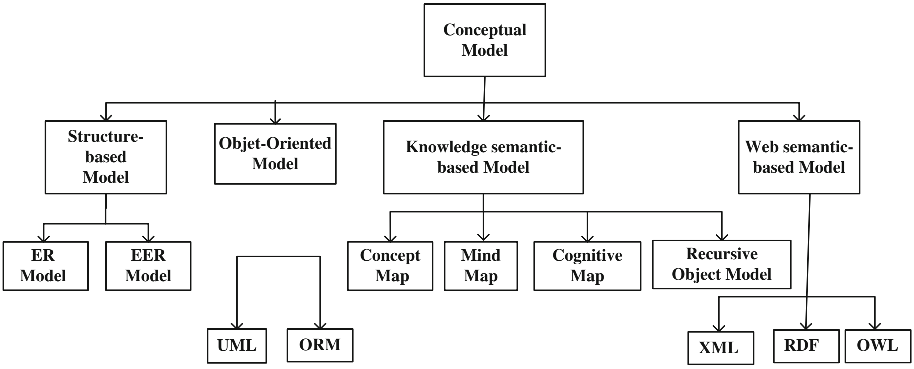

Fig. 3 Classification of the conceptual model from the functional view

as extensions to RDF(S): OIL, DAML+OIL and Web Ontology Language (OWL) (Corcho et al. 2003).

ROM: The Recursive Object Model (ROM) is a new graphic language. It was proposed by Zeng (2001); Zeng et al. (2007); Zeng (2007); Zeng et al. (2004), to represent natural language used in engineering, which is mainly composed of statements. The ROM uses only five basic symbols to represent object, compound object, constraint relationship, predicate relationship and connection relationship. Those five elements are derived mathematically and are proven sufficient for technical English through enumeration. ROM is a useful tool in analyzing the natural language, especially when extracting the main ontological words. Using the five graphic symbols, ROM can express all the cases.

Based on the description of these different conceptual models, we classified these models into four types from the functional view: structure-based model also called traditional model, object-oriented (OO) model, knowledge semantic-based model, and web semantic-based model as shown in Fig. 3.

Our classification is different from the existing classification such as classification based on expression form or based on modeling objects. The classification from the functional view is based on the function and application of the conceptual models. This kind of classification is more intuitive than the others. It carries more useful information of the conceptual models.

The first type is the structure-based model that was proposed most early, such as ER and EER mainly used for structure-based design. As the development of OO method, OO conceptual models such as UML and ORM were developed quickly mainly used for object-oriented system analysis and design. Recently Ontology has been proposed as an important and natural means of representing web resources. The third type is knowledge semantic-based model that is mainly used to represent the knowledge and knowledge processing. It includes Concept map, Mind map, Cognitive map and ROM. The fourth type related to Ontology is web semantic-based model that includes XML, RDF(S) and OWL.

We will select one or two as representative model from each type for the later comparison. Therefore the classification from the function view is the base of the comparison.

3 Criteria

Since there are many conceptual models, it is important to find out their common and different points. We will review the comparison study of conceptual models. Based on the existing comparison the criteria will be proposed. The criteria are the foundation of the comparison. It plays an important role as the metrics.

Shoval and Shiran gave an experimental comparison of design quality (Peretz Shoval 1997). They compared Entity-Relationship (ER), with its many extensions (generally termed EER) and OO (Objected-Oriented) data models from the point of view of design quality. Quality is measured in terms of (a) concreteness of the conceptual schemas being designed; (b) time to complete the design task, and (c) designers' preferences of the models. Results of an experimental comparison of the two models reveal that the EER model surpasses the OO model for designing unary and ternary relationships. Because our results support the EER model across all these dimensions strengthens the validity of the experiment. Furthermore, these results (on quality of design) are consistent with earlier results on user comprehension of EER and OO schemas, and do not contradict other results on correctness of design. Aguirre-Urreta and Marakas also gave a comparison between EER and OO (Aguirre-Urreta 2008).

Halpin and Bloesch addressed a comparison between UML and ORM (Terry Halpin 1999). Object Role Modeling (ORM) (Halpin 2001) is a powerful method for designing and querying database models at the conceptual level, where the application is described in terms easily understood by non-technical users. Halpin and Bloesch examined the relative strengths and weaknesses of ORM and UML for data modeling, and indicated how models in one notation can be translated into the other.

Moody gave theoretical and practical issues in evaluating the quality of conceptual models, and discussed the current state and future directions of the conceptual model quality (Moody 2005). Moody conducted a review of research in conceptual model quality and identifies the major theoretical and practical issues which need to be addressed. He considered how conceptual model quality frameworks can be structured shown in Table 1, how they can be developed shown in Table 2, how they can be empirically validated and how to achieve acceptance in practice. Also, he argued that the current proliferation of quality frameworks is counterproductive to the progress of the field, and that researchers and practitioners should work together to establish a common standard for conceptual model quality. Still he described some initial efforts towards developing a common standard for data model quality, which may provide a model for future standardization efforts. Moody's purpose is to address the issues about the quality frameworks of the conceptual models. It doesn't focus on the conceptual models themselves. It is very different from our work.

Teeuw and Berg introduced general quality criteria for conceptual models, which are independent of the application domain of the model (Teeuw 1997). They described an evaluation framework for behavior models used for modeling business processes, which is more specific than the general quality criteria. Besides, they introduced the conceptual framework as developed in the Testbed project, which was directed at improving business processes using a model-based approach. They argued that this conceptual framework meets the quality criteria described in the first part of the paper, and we illustrate the use of this framework by an example of a car insurance company. Teeuw proposed the following quality criteria.

Completeness: The concepts must be expressive enough to capture all ”essential aspects” (referring to the requirements of users or the market) of the real world.

Inherence (propriety): The concepts should be straight to the point and focus on essential aspects only (so no “nice features” which you ”get for free anyway”).

Clarity: A designer must be able to comprehend the concepts and rules, as well as be able to apply them in models without spending too much time and effort (subjective).

Consistency: The concepts must not conflict with each other in representation of (abstraction from) aspects of the real world. Consistency implies non-ambiguity: a concept has only

Table 1 Goals of criteria, sub-criteria and metrics

General criteria, sub-criteria and metrics		Goals
General criteria	Effectiveness	Conceptual model performs effectively
	Efficiency	Conceptual model performs efficiently
Sub-criteria	Syntactic	Syntactic correctness
	Semantic	Correctness and completeness
	Validation	Consistency, testability and applicability
Metrics for sub-criteria syntactic	Expressivity	Everything relevant may be expressed without too much effort
	Clarity	Easy to understand and use
Metrics for sub-criteria semantic	Semantic preserving	Models preserve their original meaning as much as possible
	Semantic stability	Models retain their original intent in the face of changes to the application
	Semantic relevance	Only conceptual relevant details need be modeled
Metrics for sub-criteria validation	Formal foundation	Consistency based on theory
	Empirical validation	Testability based on experiment and application fields

Table 2 Comparison of clarity

Conceptual model	Clarity
ER	Excellent
OO	Fair
Concept map	Excellent
ROM	Excellent
RDF	Good

a single meaning in the real world. The condition that there must not be two concepts with the same meaning in the real world is expressed by the criterion parsimony (restrict to essential aspects and give them a unique representation), which is considered as a special case of inherence.

Orthogonality (modularity): Independent aspects of the real world must be captured by different concepts, and (complementary) strongly related aspects should be represented by related concepts.

Generality: The concepts should be as independent as possible from any specific application or application domain.

Lindland et al. (1994) proposed three kinds of quality and a framework that not only identified major quality goals, but gave the means for achieving them. Their framework had two unique features: (1) it distinguished between goals and means by separating what they were trying to achieve in conceptual modeling from how to achieve it. They had made the goals more realistic by introducing the notion of feasibility; (2) it was closely linked to linguistic concepts because they recognized that modeling was essentially making statements in some language. Lindland proposed the followings three kinds of quality:

Semantic quality is the degree of correspondence between the conceptual model and the real world. If a model contains statements that have no correspondence in the real world, the model is invalid. In the reverse case, the model is incomplete.

Syntactic quality is the degree of correspondence between a conceptual model and its representation. The set of syntactic errors contains all statements that can be expressed in the model, but (can) not (be made) in the language.

Pragmatic quality is the degree of correspondence between the conceptual model and its (individual) interpretation, i.e., the degree up to which a model is understood.

Ter Hofstede (1993) discussed the Expressiveness in conceptual data modeling. They made extensions to an existing data modeling technique (NIAM Nijssen and Halpin 1989) which make it possible to naturally represent objects with complex structures. These extensions will be motivated from a practical point of view by examples and from a theoretical point of view by a comparison with the expressive power of formal set theory and grammar theory. Ter Hofstede (1993) addressed a general conceptual data modeling technique was introduced, the Predicator Set Model (PSM), which was capable of representing complex structures in a natural way. From a theoretical point of view the new constructs were motivated by a comparison with the axioms of formal set theory and with grammar theory. This comparison was particularly relevant for hypermedia, since document structures were usually described by means of context free grammars. Villa et al. (2009) gave a review of emerging semantic approaches to environmental modeling.

Among these comparison work, some of them are too particular such as the comparison between two models, and some are too sketch to use. We attempt to give a comparison that is different from the work existed.

Before the comparison, we should give the criteria as the base of comparison. Here we propose a criteria with hierarchy structure by using the classification given in the Sect. 2. When comparing and evaluating a conceptual model, no single criteria will be able to serve all the models. Different conceptual models are evaluated by different quality criteria. Therefore multiple criteria framework will be needed for different types of conceptual models. Base on ISO/IEC 9126 ((ISO) ISO, ISO Standard 9000-2000 2000; International Standards Organisation (ISO) IECI, ISO/IEC Standard 9126 2001), we agree that conceptual model criteria referred to as quality conceptual model should be decomposed into a hierarchy structure. It is a three-level structure namely general criteria, sub-criteria and metrics. A set of general criteria are defined at the top level that can apply to all types of conceptual models. Then these criteria will be expanded to sub-criteria in more detail for different types of conceptual models.Then these sub-criteria are expanded to metrics in further detail for particular conceptual models.

Corresponding to the classification proposed in the present paper, we give a criteria framework with a three-level structure shown in Fig. 4. The top level is general criteria that can apply to all types of concetual models including static conceptual models and dynamic conceptual models. The second level is sub-criteria that are designed for the different types of conceptual models, such as, sub-criteria for data model, sub-criteria for process model, and sub-criteria for behavior model. The third level is more detailed metrics that are used to expand these sub-criteria.

Figure 4 discribes the hierarchy structure of the criteria framework. We need further specify the criteria framework to get the definitive crerteria. Therefore general criteria, sub-criteria and metrics should be defined by using a single, concise sentence. The specific criteria are defined in Fig. 5.

To be cautious, we define two general criteria effectiveness and efficiency which can apply to all types of conceptual models. Based on the consideration that we focus on conceptual data models in the present paper, we define three sub-criteria namely syntactic, semantic and validation which can apply to conceptual data models. Each sub-criteria are expand to two or more detailed metrics.

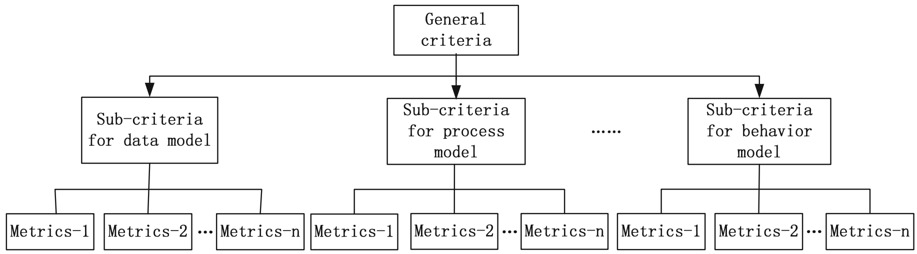

Fig. 4 The hierarchy structure of the criteria framework

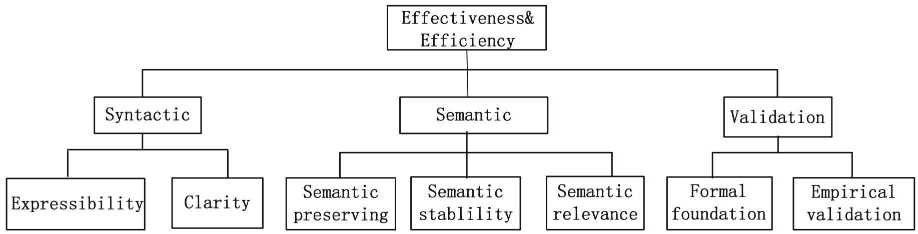

Fig. 5 The criteria for the comparative study of conceptual model

Linguists treat semiotics as consisting of three parts: syntax, semantics and pragmatics (Chen et al. 1999). Syntax defines the rules for forming sentences. Semantics is concerned with the meaning of words and sentences. Pragmatics deals with practical results, reasons and values. Here we use syntax and semantics as the first sub-criteria syntactic and the second sub-criteria semantic. The third sub-criteria are defined as validation (Dunn et al. 2005) that deals with the conformance of a given conceptual model with respect to user requirements. The third sub-criteria Validation is one of the important issues that contribute to decide whether a conceptual model is good or not (Mtais 2002). One of the frequent techniques proposed for validation is the conformance of conceptual models with applications processes. This is much more devoted for checking the completeness than to check whether conceptual entities and relationships represent effectively the semantics in users' minds.

Each sub-criteria is expanded to the corresponding metrics. The metrics are used to evaluate the conformance of conceptual models to the sub-criteria. The metrics of the sub-criteria syntactic are expressivity and clarity. The metrics of the sub-criteria semantic are semantic preserving, semantic stability and semantic relevance. The sub-criteria validation can divided into two parts. One is theory validation that evaluates whether a conceptual model has a formal foundation to achieve completeness. The other is empirical validation that evaluates completeness using experiments, case studies or application. Therefore the metrics of the third sub-criteria validation are formal foundation and empirical validation. By using the software quality characteristics, we describe the goals of these criteria, sub-criteria and metrics shown in Table 1.

4 Comparison

From each type we choose one or two conceptual models that could be called equivalence conceptual model. The equivalence conceptual model can represent the other models of the type to which the equivalence model belongs. According to the criteria given above, the comparison is conducted among these representative conceptual models instead of the exhaustive comparison among all conceptual models.

Therefore, one or two representive conceptual models need be selected for the four types of conceptual models obtained in the Sect. 2 of this paper: structure-based model, object-oriented model, knowledge semantic-based model and web semantic-based model. Accordingly ER, OO, Concept Map, ROM and RDF are selected as the equivalent models to represent these four types of conceptual models. For the reason that ROM is different from the other knowledge semantic-based models, we choose two conceptual models including Concept Map and ROM to represent the third type of conceptual models referred to as Knowledge semantic-based model.

The following comparison focus on these metrics: expressivity (Basic form also named Structure), clarity, semantic preserving, semantics stability, semantic relevance, formal foundation and empirical validation (including case study and application fields). By using these metrics we evaluate the three sub-criteria including syntactic, semantic and validation.

4.1 Expressivity

We compare the basic form of the five conceptual models shown in Fig. 6. The goal of expressivity is that everything relevant may be expressed without too much effort. The basic form of the five conceptual models mainly includes the form of concept and the relationships between the concepts.

Form the table we can conclude that all these five conceptual models can express the basic elements such as concepts and the relationships between them by using different modeling methods. Besides the basic form some conceptual models provide different methods to improve the expressivity which is beyond the scope of this paper. Generally the expressivity is proportional to the complexity. It means there is a compromise position between the expressivity and complexity.

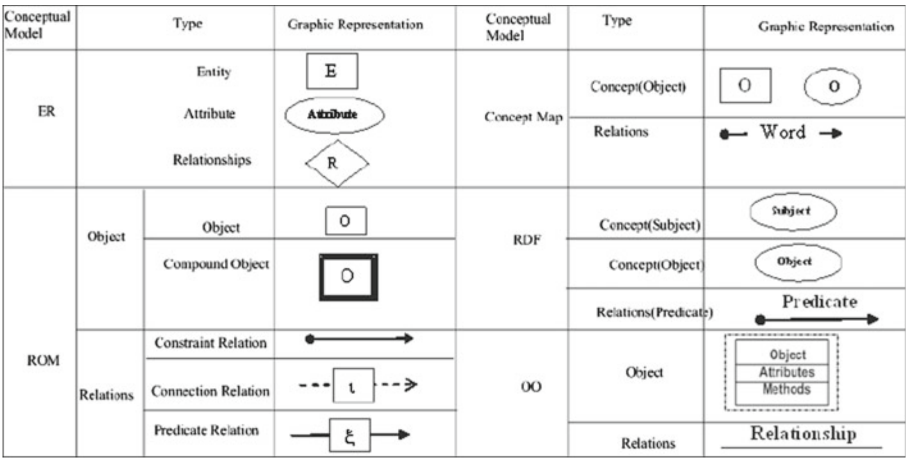

Fig. 6 The criteria for the comparative study of conceptual model

4.2 Clarity

Clarity is a measure of how easy a conceptual model is to understand and use. A conceptual model should be unambiguous. Ideally, the meaning of diagrams or textual expressions in a conceptual model should be intuitively obvious. Clarity of a conceptual model may be more important than the apparent complexity of the model when a model is used for developing domain understanding (Gemino and Wand 2005).

It is not easy to objectively evaluate Clarity. Based on the comparison of Expressivity given above and the comparison already existed, we will give a comparison of Clarity as objectively as possible.

The result of the comparison of clarity is shown in Table 2. Comparatively, ER (Scheuermann 1979), as one kind of structure-base conceptual models and traditional models, is easy to use and intuitive. OO model (Al QPe 2001) such as UML (Bonnell and Davis 2007), the use of formal specification languages provides a solid mathematical background and deducible formal properties such as soundness and completeness. However, the use of UML tends to be too complex and unfriendly. It is not easy to understand for the unprofessional users. More users thought that ER/EER model is easier to use than OO model. Concept Map visualized the relationships among different concepts. Concepts are connected with labeled arrows, in a downward-branching hierarchical structure. It is relatively easy to use and understand. RDF is particularly intended for representing metadata about Web resources. Basically RDF is easy to use and understand with a little complexity caused by using the expression form of URIs and XML. ROM uses only five basic symbols to represent object, compound object, constrain relationship, predicate relationship and connection relationship. It is easy to understand and use. Also ROM is rather intuitive. Indeed, it is difficult to completely quantify Clarity. We will further study this clarity comparison in future work.

4.3 Semantics

Semantics is the study of meanings. It includes two different meanings: internal meaning and extensional meaning. In addition to the internal meaning expressed in basic form, different conceptual models contain different semantic information referred to as extensional meaning. The semantic information of the five conceptual models is summarized shown in Table 3.

To compare the semantics of these five conceptual models, we choose the following semantic features from Maryanski (1988):

(1) Unstructured object representation is classified as limited or enhanced, depending on the degree to which the conceptual model provides nontraditional data types.
(2) Relationship representation (Storey 2005) is considered to be independent, entities (objects), tables, methods (services or functions), or attributes, depending on the manner in which the model presents the relationships to the user.
(3) Standard abstraction. The abstraction that have most frequently been identified for use in semantic data models are classification, generalization, aggregation and recursive.
(4) Network versus hierarchy. Virtually all conceptual models offer a diagrammatic construct for the conceptualization of a model. In most such models, this graph represents the fundamental modeling abstraction of the model. The nature of the graph (Network vs. hierarchy) plays an important role in the characterization of the conceptual model.
(5) Derivation/ Inheritance. By derivation/inheritance conceptual models handle repeated information.

Table 3 Semantic information of these five conceptual models

Conceptual model	Semantic information
ER	(1) Indicate the type (1:1, 1:m, or m:n) of mapping for the relationships
	(2) Provide some semantic control over the data such as naming requirements
	(3) Consider additional relevant semantic aspects. Define logical path as a sequence of connections between two entities. Shorter path carries more information than a longer one
OO	(1) Classes including attributes, methods and class relationships are defined
	(2) Use qualified associations in many cases. Provide support for sub-typing, including multiple inheritances
	(3) Add textual constraint annotations to the notation
Concept	(1) Form an answer to a focus question
map	(2) Inclusion of cross-links, helps users see how a concept in one domain of knowledge represented on the map is related to a concept in another domain shown in the map
	(3) Specific examples of events or objects helps to clarity the meaning of a given concept
ROM	(1) Be sufficient to represent a text composed of declarative sentences
	(2) Cover the basic parts of speech, sentence structures, and cohesions between words and sentences, and reserve all the information of a natural language
	(3) The idea of recursive implies the relationships between the simple object and complex object. It is flexiable to convert between them and satisfy different requirements for abstract
RDF	(1) Triple and graph
	(2) Include blank node (b-node), Literals (Data-types), Containers, Collections, Reification, Annotation, Entailment rules (rule inference)
	(3) Provide a number of additional capabilities, such as built-in types and properties

(6) Relationship semantics. Some models leave the expression of the semantics of cardinality, null values, inverse relationships, derivations, inheritance, or default values to the designers. Other models completely define the behavior of one or more of these features.
We also define one semantic metrics "Semantic reserving" and choose the other semantic metrics "Semantic stability" to conduct the semantic comparison.

(7) Semantic preserving is a measure of how well conceptual models preserve the original meaning of a natural language. That is to say, for one language unit such as text or sentence, we use different conceptual model to model it. After the operation of modeling, we will get different abstract for the language unit. So there is different degree of semantic preserving for different conceptual models. Among these five models, two of them, ER and ROM provide the translation rules between a natural language and conceptual models. It makes these two conceptual models have a good ability of semantic preserving. Semantic preserving is classified as limited or enhanced, depending on the degree to which the conceptual model preserves the original meaning of a natural language.
Actually the formal foundation of a conceptual model also improves the ability of semantic reserving. Based on the formal foundation, some extensional knowledge that is different from internal knowledge is derived by using reference. We will compare the formal foundation of conceptual models in the next section.

(8) Semantic stability (Terry Halpin 1999) is a measure of how well models retain their original intent in the face of changes to the application. Attribute-free models are more

Table 4 Comparison of semantic features and semantic metrics

	ER	OO	Concept map	ROM	RDF
Unstructured object representation	Limited	Enhanced	Enhanced	Enhanced	Enhanced
Relationship representation	Independent and tables	Independent and methods	Independent	Independent and entities	Independent
Standard abstraction	Aggregation	Aggregation generalization classification	Generalization	Recursive	Aggregation classification
Network versus hierarchy	Strong network	Hierarchy	Hierarchy	Network	Network
Derivation or inheritance	No	Inheritance	No	No	Inheritance
Relationship semantics	User selectable	Predefined	User defined	Auto preserved	Predefined
Semantic preserving	Enhanced	Limited	Limited	Enhanced	Limited
Semantic stability	Enhanced	Limited	Enhanced	Enhanced	Enhanced

stable because they are free of changes caused by attributes evolving into other constructs or vice versa. Use of attributes in the base model decreases semantic stability. Therefore, on this point to say, the attribute-related OO model is less stable than the other models. Semantic stability is classified as limited or enhanced, depending on the degree to which the conceptual model preserves the original intent in the face of changes to the application.

A semantic comparison is conducted among these five conceptual models from the eight points given above. The result is shown in Table 4.

The last semantic metric semantic relevance (Terry Halpin 1999) requires that only conceptual relevant details need be modeled. Any aspect irrelevant to the meaning should be avoided. The metric of semantic relevance should be evaluated combining with application fields. Different application fields need different level of abstraction. Sometimes more information needs to be reserved. Sometimes it is opposite. Considering this reason, it is difficult to give a uniform comparison of semantic relevance. The metric of semantics should be evaluated combining with the specific application.

4.4 Formal foundation

Formal foundation is needed to ensure unambiguity and executability, and allow formal proofs of equivalence and implication between alternative conceptual models. By using automated reasoning techniques, formal foundation can also facilitate the development of a correct implementation. Based on deep understanding of these five conceptual models, we found that some conceptual models have good formal foundation and others lack theoretical foundation. The comparison of formal foundation is shown in Table 5.

Table 5 Comparison of formal foundation

Conceptual model	Formal foundation
ER	Set theory and theory of relations
OO	Lack formal foundation
Concept map	Lack formal foundation
ROM	Set theory, mathematical relations and logic
RDF	Model theory and assertion logic

The form foundation of ER is set theory (Devlin 1993) and the theory of relations (Ulam SMaB 1990).The theory of relations is entirely grounded in set theory. The ER model views the real world as a construct of entities and association between entities. It incorporates some of the important semantic information about the real world. The model can achieve a high degree of data independence based on set theory and theory of relations. Set theory is the branch of mathematics that studies sets, which are collections of objects. The entity-relationship model adopts the more natural view that the real world consists of entities and relationships. Here entities are entity sets which are collections of entities. Relationships between entities are represented by the relations. It is corresponding to set theory and theory of relations.

OO lacks a simple model theoretical foundation for definition and discussion. Most researchers in the area of object orientation acknowledge the need for formal methods in OO. The development of concurrent object-based conceptual models has suffered from the lack of any generally accepted formal foundation for defining their semantics.

ROM is based on strong mathematical foundations. It is based on Set Theory, Mathematical Relations, and Logic. Axiomatic theory of design modeling is a logical tool for representing and reasoning about object structures (Zeng 2002). It provides a formal approach that allows for the development of design theories following logical steps based on mathematical concepts and axioms. The primitive concepts of universe, object, and relation are used in the axiomatic theory of design modeling, based on which two axioms are defined in the axiomatic theory of design modeling.

RDF is an assertion logic based on model theory, in which each triple expresses a simple proposition. An assertion is a speech act in which something is claimed to hold. This imposes a fairly strict monotonic discipline on the language, so that it cannot express closed-world assumptions, local default preferences, and several other commonly used non-monotonic constructs. Particular uses of RDF, including as a basis for more expressive languages such as DAML+OIL (2001) and OWL (Smith and Deborah 2004) based on Description Logic (Nardi and Brachman (2002)), may impose further semantic conditions and such extra semantic conditions can also be imposed on the meanings of terms in particular RDF vocabularies.

Concept Map also lacks formal foundation. It is based on psychological foundation and epistemological foundation instead of formal foundation. Early learning of concepts (Macnamara 1982) is primarily a discovery learning process. Later new concept and propositional learning is mediated heavily by language, and takes place primarily by a reception learning process. It is coming to be generally recognized now that the meaningful learning processes are the same processes used by scientists and mathematicians, or experts in any discipline, to construct new knowledge. The fundamental idea in cognitive psychology is that learning takes place by the assimilation of new concepts and propositions into existing concept and propositional frameworks held by the learner.

Table 6 Comparison of application fields

Conceptual model	Application fields
ER	Representing data object in database design
OO	System analysis and design, Data warehouses
Concept map	Organizing and representing knowledge
ROM	Representing natural language in engineering design
RDF	Representing information about resource in the world wide web or things that can be identified on the web

4.5 Application fields

Different conceptual models fit different fields. The comparison is shown in Table 6.

ER conceptual model is used to visually represent data objects. For the reason that ER model maps well to the rational model, presently it is commonly used for database design. Accordingly, Structured-based models are mainly used in engineering database design.

OO conceptual model is widely used in OO system analysis, design and data warehouses (Juan Trujillo et al. 2001). Accordingly Object-Oriented models are mainly used in engineering system analysis and design.

Concept map has subsequently been used as a tool to increase meaningful learning in the sciences and other subjects as well as to represent the expert knowledge of individuals and teams in education, government and business. It is also used to stimulate the generation of ideas, and are believed to aid creativity. Concept map can also be seen as a first step in ontology-building.

ROM is a more recent and promising conceptual model. It is mainly used to represent natural language especially in the field of engineering design. ROM retains the semantics of natural language using five basic symbols. It is sufficient and necessary to represent all the linguistic elements in technical English. So it can be applied into the application fields related to natural language processing.

In addition to organizing and representing knowledge, Knowledge semantic-based models are widely used in engineering applications, such as business processes and design decision support. Yoo et al. (2007) suggested a method to redesign business processes from the viewpoint of knowledge flows using a knowledge map. Cheah et al. (2008) constructed a manufacturing-environmental model for assembly design decision support by using fuzzy cognitive map that is one of Knowledge semantic-based models.

RDF is used to represent information about resource in the world wide web. It is particularly intended for representing metadata about Web resources. It can also be used to represent information about things that can be identified on the Web. Particular use of RDF includes as a basis for more expressive languages such as DAML+OIL and OWL. Web semantic-based models, which are Ontology-related models, are also broadly applied in engineering application (Kim et al. 2009) proposed a method that apply ontology into collaborative design. The kind of models were also used in product retrieval and reuse (Li et al. 2005), design information extraction and retrieval (Li and Ramani 2007), and collaborative product development (Kim et al. 2006). Table 13 shows the comparison results of engineering applications of different types of conceptual models.

In this section, we give the comparision according to the criteria given above. From the comparison we can see that different conceptual models have different characteristic including different forms and different semantic ability. Some of them has strong formal foundation and others do not have. It results in that different conceptual models fit different fields. The users should choose the appropriate conceptual model combining with their requirements.

4.6 A case study

We will give a comparison by using a case sentence. Here is a test sentence "The company has 50 plants located in 40 states and approximately 100,000 employees". The implementations of the case sentence are shown in Figs. 7–10 accordingly by using the conceptual models ER, ROM, Concept map, and OO.

Because RDF is mainly used to represent Web resources which has a Uniform Resource Identifiers (URIs), we make the assumption that the resources appear in the example sentence are identified in the Web.

"The company" is identified by http://www.example.org/company.

"50 plants" is identified by http://www.example.org/Plant/plants.

"40 states" is identified by http://www.example.org/State/states.

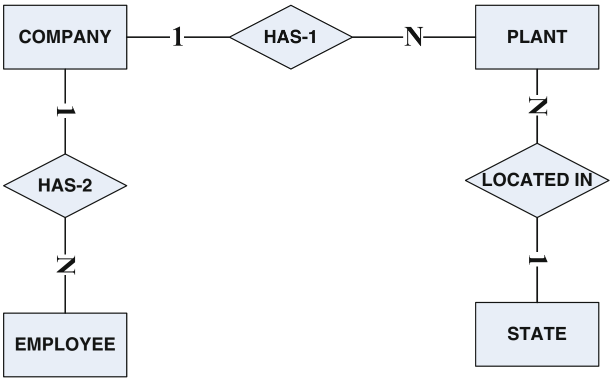

Fig. 7 ER implementation of the case sentence

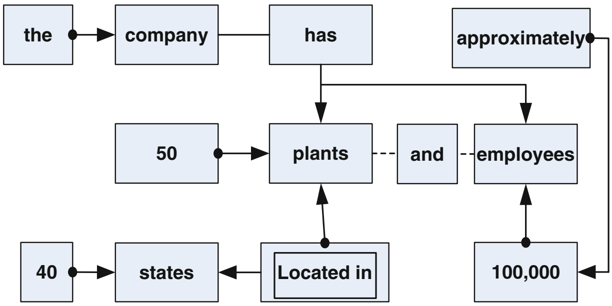

Fig. 8 ROM implementation of the case sentence

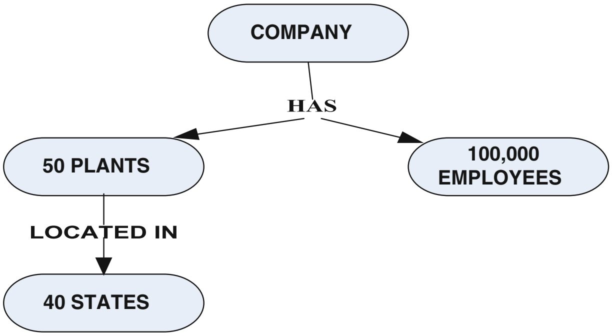

Fig. 9 Concept map implementation of the case sentence

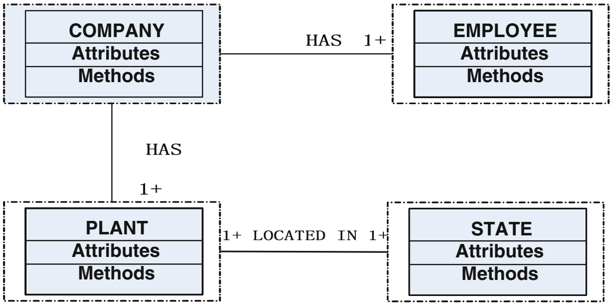

Fig. 10 OO implementation of the case sentence

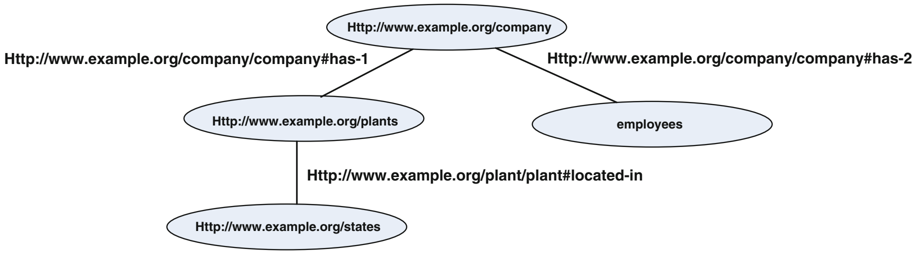

Fig. 11 RDF implementation of the case sentence

The RDF implementation of the case sentence is shown in Fig. 11.

According to the comparison above, the following conclusions can be obtained.

(1) These five models have good effectiveness.
(2) OO model has better expressivity while it has less clarity than the other four conceptual models. For the reason that better expressivity means more complexity, more complexity results in less clarity. There is a trade-off between expressivity and clarity.
(3) ER and ROM have better semantic preserving ability than the other three conceptual models. OO model has less semantic stability than the other four conceptual models.
(4) ER, ROM and RDF have good formal foundation. At the same time OO model and Concept map lack formal foundation.
(5) In the future work we need do more experiments to deeply compare the efficiency of these different conceptual models.

5 Summary

In the present paper, we focus on modeling of semantic information in engineering applications. Several popular conceptual models are introduced and their advantages and disadvantages are pointed out. Based on it the classification is given. Conceptual models are divided into four type models namely structure-based model, object-oriented model, knowledge semantic-based model, and web semantic-based model. The present paper gives the criteria of comparison and evaluation, the framework of the criteria is of hierarchy structure. It includes three levels. The top level is general criteria effectiveness" and efficiency which can apply to all types of conceptual models. The second level is sub-criteria including syntactic, semantic and validation which can apply to coceptual data models. The bottom level is metrics which is used to evaluate the sub-criteria. Based on these criteria, the comparison are conducted through the following aspect including expressivity, clarity, semantics, form validation and application fields. From the comparison users can easily get the direct understanding of the concetual models and select the right model to meet their application requirements.

Acknowledgments Dr. Yong Zeng is supported by NSERC through a Discovery Grant. Dr. Ruixuan Li is supported by National Natural Science Foundation of China under Grant No. 60873225, Dr. Kunmei Wen is thankful to the support from: Natural Science Foundation of Hubei Province (NSF-HB) under grant No. 2009CDB298, Open Foundation of State Key Laboratory of Software Engineering under Grant No. SKLSE20080718 and Innovation Fund of Huazhong University of Science and Technology under Grant No. Q2009021.

References

Aguirre-Urreta MI GMM (2008) Comparing conceptual modeling techniques: a critical review of the EER vs. OO empirical literature. The DATA BASE for Adv Inf Syst 39:9–32
Allworth S (1999) Classification structures encourage the growth of generic industry models. In: Moody DL (eds) The eighteenth international conference on conceptual modelling (industrial track). Springer, Paris, France, pp 35–46
Al QPe (2001) The OO-method approach for information systems modeling: from object-oriented conceptual modeling to automated programming. Inf Syst 26:507–534
Becker J, Rossmann M, Schutte R (1995) Guidelines of modelling (GoM). Wirtschaftsinformatik 37:435–445
Bonnell RD, Davis JP (2007) Propositional logic constraint patterns and their use in UML-based conceptual modeling and analysis. IEEE Trans Knowl Data Eng 19:427–440
Booch G, Rumbaugh J, Jacobson I (2005) The unified modeling language user guide, 2nd edn. Addison-Wesley Professional, Reading, MA
Brickley D, Guha RV (2002) RDF vocabulary description language 1.0: RDF schema W3C working draft
Buzan T, Buzan B (1996) The mind map book: how to use radiant thinking to maximize your Brain's untapped potential plume
Cheah KYK WP, Yang HJ, Kim MS, Kim JS (2008) Constructing manufacturing environmental model in Bayesian belief networks for assembly design decision support through fuzzy cognitive map. Int J Intell Inf Database Syst 2
Chen PP-S (1976) The entity-relationship model-toward a unified view of data. ACM Trans Database Syst 1:9–36
Chen PP-S (1983) English sentence structure and enity-relationship diagrams. Inf Sci 1:127–149
Chen PP-S (1997) English, Chinese, and ER diagrams. Data Knowl Eng 23:5–16
Chen PP, Thalheim B, Wong LY (1999) Conceptual modeling. LNCS 1565:287–301
Codd EF (1990) The relational model for database management, 2nd edn. Addison Wesley Publishing Company, Reading, MA
Corcho O, Ferndfndez-López M, Gmez-Prez A (2003) Methodologies, tools and languages for building ontologies. Where is their meeting point?. Data Knowl Eng 46:41–64
Date CJ (2006) Databases, types and the relational model, 3rd edn. Addison Wesley, Reading, MA

Deitel HM, Deitel PJ (2000) XML how to program, 1st edn
Devlin K, The joy of sets: fundamentals of contemporary set theory, 2nd edn
Dunn CL, Gerard GJ, Grabski SV (2005) Critical evaluation of conceptual data models. Int J Account Inf Syst 6:83–106
Fettke P, Loos P (2003) Multiperspective evaluation of reference models: towards a framework. In: Gentner GP D, Nelson HJ, Piattini M, (eds) International workshop on conceptual modeling quality evanston, IL USA
Gemino A, Wand Y (2005) Complexity and clarity in conceptual modeling: comparison of mandatory and optional properties. Data Knowl Eng 55:301–326
Gnesi S, et al (2005) An automatic tool for the analysis of natural language requirements, Int J Comput Syst Sci Eng 20:53–62
Halpin T (1995) Schema and relational database design, 2nd edn. Prentice Hall, Englewood Cliffs NJ
Halpin T (2001) Information modeling and relational databases: from conceptual analysis to logical design, 1st edn. Morgan Kaufmann, Los Altos, CA
http://www.w3.org/2001/sw/WebOnt/
(ISO) ISO, ISO Standard 9000-2000 (2000) Quality management systems: fundamentals and vocabulary
International Standards Organisation (ISO) IECI, ISO/IEC Standard 9126 (2001) Software Product Quality
Juan Trujillo MP, Gomez J, Song I-Y (2001) Designing data warehouses with OO conceptual models. IEEE Comput 34:66–75
Kanda A, et al (2008) Patent driven design: exploring the possibility of using patents to drive new design. Tools and Methods for Competitive Engineering Conference. Izmir, Turkey
Kim KY, Chin S, Kwon O, Ellis RD (2009) Ontology-based integration of morphological information of assembly joints for network-based collaborative assembly design. Artificial Intell Eng Des Anal Manuf (AI EDAM) 23:71–88
Kim KY, Manley DG, Yang HJ (2006) Ontology-based assembly design and information sharing for collaborative product development. Comput Aided Des (CAD) 38:1233–1250
Li Z, Raskin V, Ramani K (2008) Developing engineering ontology for information retrieval, Trans ASME J Comput Inf Sci Eng 8:21–33
Li Z, Anderson DC, Ramani K (2005) Ontology-based design knowledge modeling for product retrieval and reuse. 15th Int'l Conference on Engineering Design (ICED'05)
Li ZJ, Ramani K (2007) Ontology-based design information extraction and retrieval. Anal Manuf (AI EDAM) 21:137–154
Lindland OI, Sindre G, Solvberg A (1994) Understanding quality in conceptual modeling. IEEE Softw 11: 42–49
Macnamara (1982) Names for things: a study of human learning. M.I.T. Press, Cambridge, MA
Mala G SAaGVU (2006) Automatic construction of object oriented design models [UML diagrams] from natural language requirements specification. Pricai 2006: Trends in Aritificial Intell, Proceedings, pp 1155–1159
Martin J (1991) Information engineering: introduction, 1st edn. Prentice Hall, Englewood Cliffs NJ
Maryanski JPaF (1988) Semantic data models. ACM Comput Surv 20:153–189
Moody DL (2005) Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. Data Knowl Eng 55:243–276
Mtais E (2002) Enhancing information systems management with natural language processing techniques. Data Knowl Eng 41:247–272
Nardi D, Brachman RJ (2002) An introduction to description logics. Cambridge University Press, Cambridge, MA
Nijssen GM, Halpin TA (1989) Conceptual schema and relational database design a fact: oriented approach. Prentice-Hall, Englewood Cliffs NJ
Novak JD, Canas AJ (2008) The theory underlying concept maps and how to construct and use them. Florida Inst Hum Mach Cogn
OMG (2007) OMG unified modeling language (OMG UML), Infrastructure, V2.1.2
OMG (2008) Introduction to OMG's unified modeling language
Peretz Shoval SS (1997) Entity-relationship and object-oriented data modeling-an experimental comparison of design quality. Data Knowl Eng 21:297–315
Peterson JL (1981) Petri net theory and the modeling of systems. Prentice Hall PTR, Englewood Cliffs NJ
Petri CA (1962) Kommunikation mit automaten. University of Bonn, West Germany
Reisig W (1985) Petri nets, an introduction. Springer, Berlin
Reisig W (1992) A Primer in Petri net design. Springer, Berlin
Rumbaugh J, Jacobson I, Booch G (2004) The unified modeling language reference manual, 2nd edn. Addison-Wesley Professional, Reading, MA

Scheuermann GSaP (1979) Multiple views and abstractions with and extended entity relationship model. Comput Lang 4:139–154
Smith MKCW, DL McGuinness (2004) W3C, OWL web ontology language guide
Storey VC (2005) Comparing relationships in conceptual modeling: mapping to semantic classifications. IEEE Trans Knowl Data Eng 17:1478–1489
Sven Hartmann SL (2007) English sentence structures and EER modeling. In: Proceedings of the fourth Asia-Pacific conference on conceptual modelling, pp 27–35
Teeuw WB HvdB (1997) On the quality of conceptual models
Ter Hofstede TPvdW AHM (1993) Expressiveness in conceptual data modelling. Data Knowl Eng 10:65–100
Terry Halpin AB (1999) Data modeling in UML and ORM: a comparison. J Database Manag 10:4–13
Tolman EC (2000) Cognitive maps in rats and man. Psychol Rev 55:189–208
Ulam SMaB AR (1990) On the theory of relational structures and schemata for parallel computation. University of California Press, Berkeley, CA
Villa F, Athanasiadis IN, Rizzoli AE (2009) Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. Environ Model Softw 24:577–587
W3C (2004) Resource description framework (RDF): concepts and abstract syntax
W3C (2004) RDF/XML syntax specification (Revised)
W3C (2006) Extensible markup language (XML) 1.0
Yoo K, Suh E, Kim KY (2007) Knowledge flow-based business process redesign: applying a knowledge map to redesign a business process. J Knowl Manag 11:104–125
Zeng Y (2001) Axiomatic approach to the modeling of product conceptual design processes using set theory. Department of Mechanical and Manufacturing Engineering. University of calgary, Calgary, Alberta, Canada
Zeng Y (2002) Axiomatic theory of design modeling, Transaction of SDPS. J Integr Des Process Sci 6:1–28
Zeng Y, Chen L, Wang M (2007) Automatic generation and layout of ROM diagram from English text. In: University DtC, editor. Patent Application. Canada
Zeng Y (2007) Recursive object model (ROM)—Modeling of linguistic information in engineering design computers in industry
Zeng Y, Pardasani A, Antunes H, Li Z, Dickinson J, Gupta V et al (2004) Mathematical foundation for modeling conceptual design sketches. Transactions of the ASME: J Comput Inf Sci Eng 4: 150–159