Inf Technol Manage (2007) 8:241–252
DOI 10.1007/s10799-007-0019-5

Ontology learning: state of the art and open issues

Lina Zhou

Published online: 24 March 2007

Abstract Ontology is one of the fundamental cornerstones of the semantic Web. The pervasive use of ontologies in information sharing and knowledge management calls for efficient and effective approaches to ontology development. Ontology learning, which seeks to discover ontological knowledge from various forms of data automatically or semi-automatically, can overcome the bottleneck of ontology acquisition in ontology development. Despite the significant progress in ontology learning research over the past decade, there remain a number of open problems in this field. This paper provides a comprehensive review and discussion of major issues, challenges, and opportunities in ontology learning. We propose a new learning-oriented model for ontology development and a framework for ontology learning. Moreover, we identify and discuss important dimensions for classifying ontology learning approaches and techniques. In light of the impact of domain on choosing ontology learning approaches, we summarize domain characteristics that can facilitate future ontology learning effort. The paper offers a road map and a variety of insights about this fast-growing field.

Keywords Ontology · Ontology learning · Learning techniques · Framework

The Web is evolving from a huge information and communication space into a massive knowledge and service repository. One of the enablers of the above change is ontology, commonly referred to as the conceptualization of a domain [1]. Ontology provides a sound semantic ground of machine-understandable description of digital content. It is ubiquitous in information systems [2] by annotating documents with meta-data, improving the performance of information retrieval and reasoning, and making data between different applications interoperable [3–7]. In addition, ontology-type semantic description of behaviors and services allow software agents in a multi-agent system to better coordinate themselves [8–10]. Therefore, ontology development will have a profound impact on a wide range of enterprise applications and knowledge integration.

The potential of ontologies would not be materialized until they become widely available. To date, ontology development remains labor-intensive and time-consuming. One of the major challenges lies in ontology acquisition [11, 12]. Ontology learning refers to the automatic discovery and creation of ontological knowledge using machine learning techniques. Compared with manually crafting ontologies, ontology learning is able to not only discover ontological knowledge at a larger scale and a faster pace, but also mitigate human-introduced biases and inconsistencies. Moreover, ontology learning can support refining and expanding existing ontologies by incorporating new knowledge.

Making ontologies operational in the context of the Web and other large distributed systems requires a considerable amount of research effort towards developing methodologies and technologies for constructing and maintaining domain-specific ontologies in a dynamic environment. Driven by the great potential of ontology learning in ontology development, we propose a model of rapid ontology development (ROD) and a framework for ontology learning to support ROD. The remainder of the paper is organized as follows. After discussing major research issues involved in ontology development, we present a

L. Zhou (✉)
Department of Information Systems, UMBC, 1000 Hilltop Circle, Baltimore, MD 21250, USA
e-mail: zhoul@umbc.edu

learning-oriented model for ontology development in Sect. 1, and propose a framework for ontology learning in Sect. 2. Then, we introduce and compare ontology learning approaches and techniques in Sect. 3. Next, we discuss the importance of domain while choosing ontology learning approaches, and provide recommendations for aligning domains to ontology learning approaches in Sect. 4. We identify a number of open issues in ontology learning and suggest potential solutions in Sect. 5, and conclude the paper in Sect. 6.

1 A learning-oriented model of ontology development

Ontologies are aimed to provide knowledge about specific domains that are understandable by both developers and computers. In particular, ontologies enumerate domain concepts and relationships among the concepts [13]. They may also explicitly define properties, functions, constraints, and axioms [14]. In this section, we first review major issues involved in ontology development, including ontology representation, ontology acquisition, ontology evaluation, and ontology maintenance. A further examination of the above issues demonstrates great potential of ontology learning in ontology development. To provide a systematic guidance, we introduce a learning-oriented model of ontology development. Inspired by the merits of existing tools in support of ontology development, we specify detailed requirements for integrated tools to support the proposed model.

1.1 Major issues in ontology development

Ontology representation is the most fundamental issue in ontology development. In addition to making ontologies understandable by computers and humans, an ontology representation language should also provide representation adequacy and inference efficiency. The standardization of ontology representation languages (e.g., RDF, RDFS [15], and OWL [16]) has taken big strides in the past few years. The above languages have mainly adopted a frame-based knowledge representation paradigm [17], some of which (e.g., OWL) incorporate description logics to enhance the expressiveness of reasoning systems.

Ontology acquisition refers to the creation of the content of ontologies such as concepts and relations. Given the strong dependence on domain knowledge, ontology modeling is traditionally carried out by knowledge engineers and/or domain experts. As domain knowledge evolves rapidly and workforces become increasingly distributed, subject-matter experts are not easily accessible and the experts' knowledge is likely to be incomplete, subjective, and even outdated. To keep up with the requirements of practice, people turn to other sources such as dictionaries, Web documents, and database schemas for the content of ontologies [18]. As a result, ontology acquisition can significantly benefit from ontology learning.

Ontology evaluation is another major issue as ontologies become available. Ontology evaluation can enhance the quality of ontologies, improve the inter-operability among systems, and further increase the wide adoption of ontologies. Ontologies can be evaluated from a variety of perspectives, ranging from content to technology, methodology, and application [20], using objective measures such as completeness, consistency, and correctness [19]. Among others, ontology learning is regarded as an alternative method to evaluate the content of ontologies [20].

Ontology maintenance pertains to how to organize, search, and update existing ontologies. This issue looms large as more and more ontologies are accumulated. The constant evolving of the environment of ontologies makes it very important for ontologies to be evaluated and maintained [21] to keep up with the change. SWOOGLE [22] is a crawler-based indexing and retrieval system for the semantic Web, housing several millions of Web documents in RDF and OWL. The enormous number of semantic Web pages and ontologies makes manual ontology maintenance a daunting task. As a result, automated solutions should be explored for ontology integration and mapping. Ontology learning appears to be an attractive approach to this goal.

The above discussion demonstrates that many issues involved in ontology development can benefit from ontology learning. Ontology learning enables the acquisition of ontological knowledge with little human intervention. Instead of fully relying on subject-matter experts, ontology learning exploits various forms of data within a domain, and applies machine learning (and natural language processing) techniques to discover ontological knowledge (semi-) automatically. Ontology learning not only improves the efficiency of ontology development process but also enables discovery of new knowledge by tapping into data repositories. In the next section, we present a learning-oriented model of ontology development.

1.2 Rapid ontology development (ROD): a learning-oriented model of ontology development

Ontology development is still faced with a number of challenges such as knowledge acquisition and lack of sufficiently validated and generalized development methodologies [11, 23]. With the advancement of ontology learning and ontology modeling tools, the process of ontology development has undergone rapid changes. After observing repeated ontology development processes that

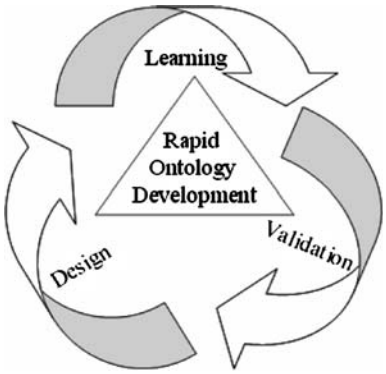


Fig. 1 Rapid ontology development (ROD)

involve ontology learning, we develop a learning-oriented model of ontology development, called ROD (Rapid Ontology Development), as shown in Fig. 1.

ROD consists of three phases: design, learning, and validation. The design phase involves the identification and detailed analysis of domains, requirements, and relevant resources with the help of users and/or domain experts. The output includes specifications of domains, intended applications of ontologies, and authoritative domain sources. Domain sources are in a variety of forms ranging from structured types such as thesaurus and glossary, to semi-structured types such as book indexes and Web pages, and free-formed types such as expert knowledge. They involve different media types such as text, audio, and image. A variety of techniques can be employed in domain analysis, such as interviews, questionnaires, and informal text analysis [24]. The understanding of user requirements and target applications helps refine those approaches.

In the learning phase, appropriate learning techniques are selected, implemented, and then applied to discover ontologies from domain sources. Ontology learning will be discussed extensively in the rest of the paper.

The learning results are evaluated and refined during the validation phase, where discovered ontologies are checked for redundancy, conflict and/or missing information. The active involvement of users and domain experts is highly desired in this phase. Once problems are identified, causes will be analyzed, remedy solutions will be developed and applied accordingly, and ontology development proceeds to the next cycle. ROD is an iterative process, which is repeated until the outcome is acceptable to users and/or knowledge engineers.

Compared with conventional ontology development methodologies, ROD can not only reduce the cost but also shorten the lifecycle of ontology development, which is crucial to semantic Web applications. Moreover, ROD fulfills quality requirements by involving users and/or domain experts early in the development process and by incorporating necessary changes in a timely fashion. Rapid construction of ontologies helps users better understand the requirements and application of ontologies. Furthermore, ROD enables ontology-based business process management and dynamic application integration to maintain competitive business advantages. Theoretically, ROD, built on top of recurring ontology development constructs and ontology learning techniques, not only allows the reuse of previously tested solutions to ontology development but also provides reusable ontologies that can be seamlessly incorporated into new applications.

1.3 Requirements for integrated tools in support of ROD

ROD could work with some emerging tools and integrated ontology development environments such as Text-to-Onto [23], ASIUM [25], Mo'k Workbench [26], OntoLT2 [27], and OntoLearn and Consys [12, 28]. To realize the vision of ROD, a host of functions need to be supported by ontology tools and technologies. Besides ontology learning, some functions that can facilitate ontology development are listed as follows:

a) Knowledge elicitation: captures domain knowledge from users directly.
b) Ontology retrieval: allows users to search ontologies based on terms, relations, properties, and so on.
c) Ontology editing: supports users in creating and updating ontologies.
d) Ontology validation: facilitates users in verifying the quality of ontologies. It may be partially automated by incorporating inference mechanism into the system. For example, description logic provides strong internal support for inference. Thus, it should be able to add additional axioms and deductive rules to ontologies [29].
e) Collaborative development: supports a team of distributed users in collaborative ontology development. Ontology development is a knowledge-intensive task. A single user may not be sufficient to cover the knowledge of a target domain. Instead, a group of users with complementary expertise should be allowed to collaborate in developing ontologies, each of whom focuses on the part of an ontology falling within the scope of his/her expertise. The consistency and integrity of ontologies can be ensured with concurrency control and security mechanisms. Some systems have emerged to support distributed collaboration in ontology development [30–32].
f) Ontology transformation and presentation: allows users to import and export ontologies encoded in different standard representation languages and to interoperate ontologies developed with different tools and

applications. Visualization is a promising approach to ontology presentation. For example, ontologies can be displayed in multiple views (e.g., tree, tab panel, and graph) at different granularities (e.g., terms, single terms, and properties) [32]. Recent versions of Protégé-2000 [33] provide graphical widgets to support users in constructing ontologies.

Each of the above functions supports one or more phases of ROD. For example, (a) supports the design phase, (b–d) supports the validation phase, and (e, f) support both design and validation phases. The functions in support of ontology learning will be discussed in detail in the next section.

2 A framework for ontology learning

We propose a framework for ontology learning in its most general form based on the related literature (e.g., [12, 23, 28]) and our experience in ontology learning (e.g., [34]). We integrate components that are commonly involved in ontology learning into the framework, which utilizes domain resources in generating ontologies.

The framework for ontology learning consists of information extraction, ontology discovery, and ontology organization. The design phase of ROD is instrumental to ontology learning. For example, domain analysis helps identify related domain sources, which can be used to discover ontological knowledge. In some cases, top-level ontologies are agreed upon during the design phase, which serve as valuable knowledge sources to guide the entire ontology learning effort.

Information extraction. A variety of data can be exploited in ontology learning, including textual documents (e.g., Web pages, papers, reports, book chapters, newspaper articles), structural data (e.g., Web site structure and book subject indexes), and usage data (e.g., the log of user navigation and search queries). Information extraction pre-processes and recognizes information in a variety of forms and converts them into the forms that can be used for ontology discovery. In particular, text documents are processed via content analysis [35] by employing a variety of natural language processing techniques, ranging from tokenization, to part-of-speech tagging, phrase structure and/or grammatical function parsing, semantic and discourse analyses [36]. Server logs of user-system interaction history need to be cleaned by identifying and extracting users, sessions, pageviews, and so on. Moreover, the content, structure, and usage data can be used in combination to support information extraction. For example, the content of pages that users frequently access consecutively or those pages that are linked together are likely to be related in content.

Ontology discovery. Supervised and unsupervised learning algorithms have been applied to discover ontological concepts and relations from the extracted information. During concept learning, words/phrases that only perform grammatical functions and words that are unlikely to carry domain-specific meanings are filtered out using information retrieval techniques [37]. The candidates for domain concepts can be sorted in the descending order of the strength of their relationships with other candidates. Relation learning generally relies on co-occurrence statistics of information [38, 39]. Approaches to relation learning vary in terms of the scope of co-occurrence, the metrics for the significance of co-occurrence, the criteria for selecting candidate concepts, and the thresholds for extracting potential relations. Some learning approaches require the assistance of domain-specific resources such as thesauri.

Ontology organization. Given the large number of possible ontological concepts and relations extracted from the learning process, an issue arises as to how to improve the usability of the discovered knowledge. Ontology organization seeks to achieve the above goal via the following steps:

- Clustering synonymous terms and their relations. Synonyms are merged to reduce redundancies. For example, given synonyms A and B, a relationship between A and C, and a relationship between B and C, we group A and B as a single concept with the corresponding relationship with C.
- Deriving inverse relations. We can derive inverse relationships from the ones that have been discovered. For example, if concept A is found to be a part of concept B, we can infer that B has a part of A.
- Discovering local centers of concepts. The strengths of the relationships between concepts vary greatly. A concept island is a group of concepts that are closely related among themselves but marginally related to concepts outside the group. Give a concept islands, we can infer the importance of concepts by identifying local centers via social network analysis. If we connect any two concepts i and j that are directly related within an island via path p_{ij}, a network will be constructed. The length of p is the total number of concepts on p minus 1. If a pair of concepts is related transitively, the length of p is greater than 1. Island center $c*$ can be found by looking for the concept(s) that has the shortest total path connecting to the rest of the concepts on the same island (the total is M), as shown in Formula (1).

$$c* = \arg\min_{c=1\ldots M} \sum_{\substack{i=1 \\ i\neq c}}^{M} p_{ci} \qquad (1)$$

- Building higher-level ontologies. The local centers discovered in the previous step can serve as the top-level ontology of the target domain. The process for identifying local centers in a concept island is repeated to find concepts for a lower-level ontology by selecting concepts that have the next shortest length of path. This process continues to the desired level where concepts constitute the leaf nodes of an ontology hierarchy. It is possible to construct such a hierarchy for each relation separately.

It is desirable to automate all the components in the framework for ontology learning by developing techniques. However, this remains infeasible and some components are carried out manually. For example, Zhou et al. [34] first crafted the top-level ontology manually based on existing domain resources, then learned ontological concepts and relations automatically, and finally mapped learning results to the top-level ontology by hand. Riloff and Shepherd [40] initiated the learning process with a set of manually selected seed concepts, which were then bootstrapped gradually. We summarize and analyze ontology learning approaches in the next section.

3 Ontology learning approaches

3.1 Classification of ontology learning approaches

A variety of approaches have been applied to ontology learning. Based on an analysis of factors that have impact on ontology learning process and/or outcome in the extant literature, we identify important dimensions for the classification of ontology learning approaches, including units to be learned, learning targets, learning strategies, and learning techniques, which are listed in Table 1. It is noted that each dimension is related to one or more components in the framework for ontology learning (See Fig. 2). For example, data sources and learning targets are related to the input and output of ontology learning respectively. Learning strategies involve both ontology discovery and ontology organization.

A learning unit can be either a multi-word phrase or a single word. In addition to concepts and relations, learning targets could be definitions that describe the concepts and axioms that constrain the interpretation of concepts and relations. Ontologies can be learned from static sources such as text documents, Web, and dictionaries as well as dynamic sources such as user-system interaction behavior and Web logs. Once the domain and data sources are identified, the decision on learning strategies can be made. The bottom-up strategy starts with text documents and gradually derives top-level ontologies while the top-down strategy begins with the top-level ontology gradually bootstraps low-level ontologies. The learning technique is the core of an ontology learning approach, which will be discussed in detail in the next section. Some ontology learning approaches depend more strongly on external knowledge than others.

Table 1 Classification of approaches to ontology learning

Dimensions	Categories
Learning units	Word and term (single and multi-word)
Learning targets	Concept, relation, definition (description of ontological concept), and axiom
Data sources	Document collection [41], Web [32, 42], dictionary [43, 44], and user interaction [45]
Learning strategies	Bottom-up [46, 47], top-down [40], and hybrid [32]
Learning techniques	Statistics-based [48], rule-based [49], and hybrid [50]
Knowledge support	Knowledge-rich [51] and knowledge-lean [52]

3.2 A taxonomy of ontology learning techniques

Learning techniques play an important role in an ontology learning approach. Ontology learning commonly involves machine learning techniques in discovering ontological knowledge. Machine learning represents a generic type of techniques that deal with extracting inductive patterns from data or building models to fit the data [53]. An ontology learning technique usually starts with formalizing data with a set of features indicative of ontological knowledge, then generates internal representations of candidate models from the formal input of data, and finally applies the best models selected based on certain criteria to produce ontological knowledge.

Drawing upon the extant literature, we propose a classification schema of ontology learning techniques, as shown in Fig. 3. The top-level consists of three categories: statistics-based, rule-based, and hybrid techniques. The majority of ontology learning techniques are unsupervised because training data annotated with ontological knowledge are commonly not available. As a result, statistical techniques have been often applied in ontology learning. A statistical model is typically represented as a probabilistic network that indicates the probabilistic dependencies between random variables [54, 55]. The statistical information computed from observed frequencies or joint distributions of the terms is used to determine concepts and their relations. Different approaches vary in how this probabilistic network is generated and which method is applied to combine individual distributions. Maximum Likelihood Estimation and Bayesian approaches are typical examples. Rule-based approaches require matching to

Fig. 2 A framework for ontology learning

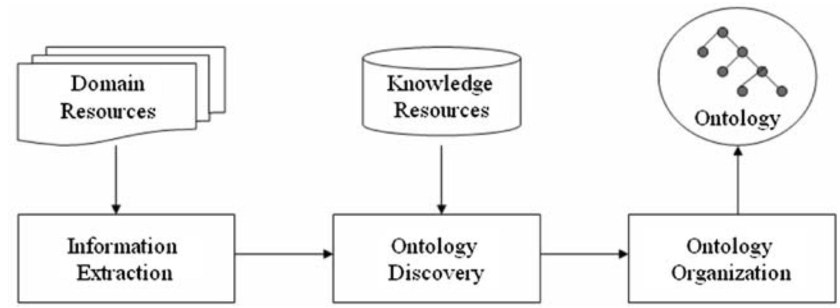
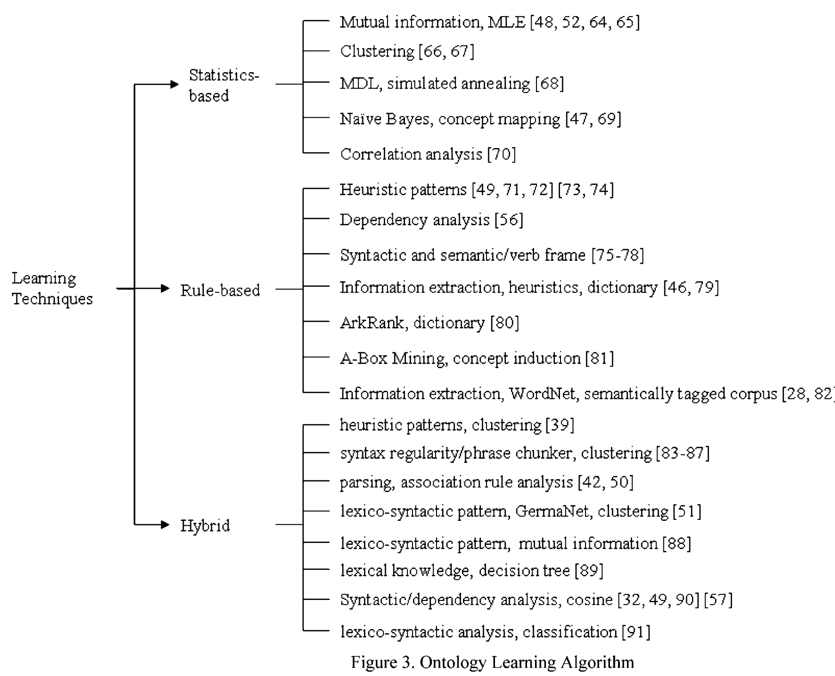


Fig. 3 Ontology learning algorithm



pre-defined rules or heuristic patterns in order to extract terms and relations. A rule-based model is typically represented as a set of rules consisting of condition testing and action execution, such as dependency relation analysis [56, 57] and anaphoric resolution [58, 59]. Hybrid approaches leverage the strengths of both statistics-based and rule-based approaches.

Among various machine learning applications, ontology learning focuses on association learning. Measures of associations can be generally defined in two ways. One is based on extrinsic features in the context, typically co-occurrences of terms, which is referred to as context similarity. The other is based on intrinsic features of terms such as their own properties and relations, which is called self similarity. There is another rarely used approach (e.g., three-layer feed forward neural network [60]), which is not discussed in this paper. Each of the learning techniques focuses on a different aspect of ontology learning and has many variants, making the selection of the most appropriate technique difficult. For example, the scope of co-occurrence analyses ranges from sequences of words (n-grams), to syntactic structures (e.g., verb-object), and to windows of fixed context. It is possible to design a hybrid measure that takes advantage of both context and self-similarities such that a measure of association may assess the context similarity of those terms that satisfy self-similarity requirements.

The analysis of ontology learning techniques reveals a trend of enhancing the performance of ontology learning with additional knowledge support such as dictionary and thesaurus. Given the lack of established resources in some domains, natural language processing techniques are introduced into ontology learning to generate linguistic information (e.g., phrase structures and dependency relationship) from the text automatically.

Like other machine learning applications, one major problem facing ontology learning is how to estimate the probability of co-occurrences that are not observed in a training corpus. Such a data-sparseness problem can be addressed with two types of approaches: smoothing and class-based methods. Smoothing methods estimate the probability of unobserved co-occurrences using observed frequency information [61, 62]. Class based models [63] distinguish between unobserved co-occurrences using classes of ''similar'' words. The probability of a specific co-occurrence is determined using generalized parameters about the probability of class co-occurrences.

Ontology learning techniques have been applied to a variety of domains, which could have impact on the selection of ontology learning techniques.

4 A classification scheme of domains

The quality of ontology learning results is usually compromised when a developer fails to recognize that a favored strategy for learning is actually uninformative and/or a specific learning approach is applied to an unsuitable domain. Poor quality often results from a mismatch between learning strategies and domains. Given the importance of matching domains with methods to reuse [92], the selection of ontology learning approaches should take the characteristics of domains into consideration. Therefore, we classify domains from multiple perspectives and make recommendations of ontology learning approaches.

Established vs. under-developed domains. With the proliferation of ontology use, more and more domain ontologies are developed. At the minimum, standard terms, relations, and even thesauri are available in those domains. Such resources make top-down and knowledge-rich approaches very attractive for ontology learning. For example, in the biology domain, it is easy to find well-defined vocabulary and even ontologies for genes and molecules. Nonetheless, many other domains remain to be under-developed (e.g., community development) and lacking of domain-specific resources. Constrained by resource availability in those domains, knowledge-lean and bottom-up approaches should be preferred in ontology learning.

Emerging vs. conventional domains. New domains are constantly emerging, driven by scientific and technological advances or national needs. Such domains (e.g., nanotechnology) generally do not yet have ontologies. The access to authoritative information sources is also likely to be limited. Thus, the bottom-up approach appears to be a good alternative for ontology learning, which can facilitate discovering new domain knowledge. Conversely, in a conventional domain, an enormous amount of information has been accumulated over time. For example, with over three hundred types of dills in gardening, the number of relevant information sources is overwhelming. Thus, it is better to start with top-level concepts and to expand them using a top-down or hybrid approach in developing ontologies.

Technology-heavy vs. technology-light domains. Technologies are the driving force and the enabler of information sharing. As a result, knowledge in technology-heavy domains is being updated at a much faster pace than technology-light ones. To this end, a hybrid approach that allows for incremental knowledge update seems promising to technology-heavy domains.

Self-contained vs. interdisciplinary domains. Unlike self-contained domains, interdisciplinary domains are prone to ambiguities associated with term usages. For example, *ontology* in philosophy has different meanings from that in artificial intelligence. As a result, a bottom-up approach may not converge to higher-level ontologies in interdisciplinary domains. A better alternative would be starting with defining the top-level ontology and then bootstrap it to acquire a full ontology.

5 Open issues in ontology learning

Tremendous progress has been made in ontology learning, owing to the contribution of a variety of disciplines such as formal modeling, computational linguistics, knowledge engineering, machine learning, and Web intelligence. Nonetheless, the ultimate goal of developing a set of full-fledged ontology learning tools that can produce high-quality learning outcome is still far from reaching. A number of challenges need to be addressed before this field matures. Some open issues in ontology learning that are worthy of future research are summarized and potential solutions are suggested in the rest of this section.

Human understandable vs. machine-understandable. Ontologies are designed to facilitate knowledge sharing and reuse between humans and software agents. In terms of the level of abstraction, ontology representation languages form a spectrum ranging from natural language at one end to formal language at the other. The two extreme ends best satisfy the needs of humans and software agents respectively. In order to bridge the gap between them, we may allow humans to access ontologies using natural language by mapping natural language words/phrases to low-level ontology concepts automatically while keeping the top-level ontology abstract.

Learning specific relations. Previous effort on ontology learning has focused on discovering generic relatedness or associations rather than specific types of relations. Among different types of specific relations, the taxonomic relation has been discussed most frequently. Other types of relations, although much less studied, are also important to some domains. For example, *part-whole* relation is critical to Sequence Ontology [93] for genomic annotation and *related_synonym* is part of Gene Ontology. Some efforts have been taken to discover *part-whole* relationship from the text [71, 94] and in learning instances of relations from the Web [95]. Additional research work is needed in discovering specific relations by ontology learning.

Learning higher-degree relations. All the relations and/or associations that have been discussed so far are binary, which relates two concepts with each other. However, higher-degree relations are required for some domains. For example, the *trust* relation describes who trusts whom on what. To our best knowledge, little work has been done in learning higher-degree ontological relations. This will be an interesting issue for further investigation, possibly by extending the approaches to learning binary relations.

Learning definitions. In view of the semantic ambiguity in interpreting a concept, the definition can play a part in enriching ontological concepts and facilitating consistent interpretation and application of ontologies. Definition learning may take on various forms: identifying and extracting new definitions, selecting from alternative definitions, and compiling definitions from pieces of information. Heuristic indicators and/or patterns such as *is defined as* and *is referred to as* have been applied in learning definitions (e.g., [96]). Moreover, the large body of literature on word sense disambiguation [75, 97–102] is conducive to selecting among alternative definitions. There have been some encouraging results on interpreting new concepts based on the definitions of component words from existing resources [28] and in learning definitions through relations [103]. Nonetheless, the effectiveness of the above approaches needs to be evaluated formally.

Term filtering. Learning ontologies from the text is subject to high false alarms caused by a large number of extraneous terms in discovering domain-specific concepts. The noise terms would engender even more superfluous associations and computation cost. Therefore, it is important to filter such terms as early as possible in ontology learning. Mutual information and traditional term weighting techniques developed for information retrieval have been utilized in selecting domain-specific terms [32, 104]. Contrast analysis is another practical means to distinguish concepts of the target domain from those of other domains [28, 67]. Nonetheless, how to select contrast domain(s) remains an issue. Such domains are expected to be diverse enough to identify domain-specific concepts and similar enough to screen out common terms. Furthermore, anaphora resolution [105] and co-reference resolution [106], aiming to identify entities referred by pronouns and definite noun phrases, has great potential in refining the discovery of concepts and relations.

Mapping to high-level ontology. How to map ontology learning results to high-level ontologies or how to construct high-level ontologies from fine-grained learning results is crucial to effectively organizing ontologies. Linking linguistic entities to ontology concepts and relations can be carried out either manually [107] or automatically [108]. For example, using path analysis (See Sect. 2), a concept that has a shorter path to other concepts can be placed higher in an ontology hierarchy. Another alternative is to treat the top-level concepts in an ontology as seeds and then to induce other related concepts in a bootstrapping manner until the desired depth is reached [40, 109]. The mapping problem becomes more complex when multiple ontologies are involved. Ontology merging or integration [110–112] remains a challenging task.

Evaluation benchmark. Human validation is still mandatory in the state of practice in ontology learning. In particular, ontology learning results have mainly been evaluated by domain experts manually. Due to limited access to domain experts and extensive effort involved in manual evaluation, it is highly desired to have benchmark corpora, which are assembled from domains of interest and annotated with corresponding ontologies, to support automatic validation. Moreover, the benchmark corpora allow evaluating both usefulness and accuracy of resulting ontologies quantitatively. For example, the quality of an ontology can be measured by the degree to which the ontology ''fits'' a corpus [113], which can also be viewed as a measure of validity of the ontology to the domain. Without the common benchmark corpora, it is difficult to compare different approaches on an equal basis. The benchmark represents a shared understanding of the task at hand, allowing one to identify the strengths and weaknesses of an ontology learning approach objectively and to concentrate on improving the state of ontology learning. Furthermore, given the corpora annotated with ontological knowledge, supervised approaches (e.g., [74]) to ontology learning can also be pursued.

Incremental ontology learning. As domain knowledge evolves over time, ontologies need periodic updates, which may become the central task of ontology maintenance. It would be more effective to reuse existing ontologies instead of re-learning the entire ontology from scratch, especially as domain ontologies proliferate. An incremental learning approach [114] that can accommodate new changes by updating existing ontologies incrementally seems very attractive.

Levels of ontology learning. According to the level of details, ontologies are classified into four types: meta-level, reference, shareable, and domain ontologies [115]. Loosely defined, ontologies span the spectrum with formal ontologies at one end and knowledge base at the other end. Some researchers [116] argue that ontologies should be abstract and stay only at the top level, while others think that ontologies are similar to terminological knowledge base (e.g., [117]). In the latter case, ontologies are populated with terms that manifest the concepts, relations, and constraints.

Generally, ontology learning can directly benefit creating low-level ontologies. Another decision involved in ontology learning is whether to learn concepts or instances. There are no golden rules to follow in making the distinction. This is largely dependent upon the scope of ontology learning. For example, ''Dell Notebook'' can be considered as either a subclass or an instance of ''Notebook''. It is also possible to learn new instances for pre-specified ontology classes and relations [95].

Multi-agent learning. In multi-agent learning, a group of distributed, learning agents improves its group performance through collective experience. The agent-based paradigm is ideal for supporting distributed and collaborative ontology learning. Agents are able to locate and translate disparately referenced concepts and improve concept precisions as they enhance their experience from others. It is proven that agents are able to: 1) learn diverse ontologies; 2) locate, share, and integrate knowledge; 3) improve group performance through experience; and 4) introduce novel approach and novel algorithms [118–121].

Learning beyond text. Text is probably the most widely available medium for learning ontologies. Ontology learning can also leverage multimedia information such as audio, video, and image by employing multimedia technologies. For example, content-based image retrieval techniques can be applied in learning ontologies from image content. In addition, it has also been suggested to draw hypotheses about term relations on the basis of user observations [122, 123]. For example, if a user of a retrieval system combines two terms by OR in his/her query (and further requirements hold), these terms are probably synonyms.

6 Conclusion

The potential of ontologies in supporting knowledge sharing and reuse has attracted wide attention from researchers and practitioners across many disciplines. However, making ontologies operational within the Web and other distributed systems still requires a considerable amount of research and development effort to construct and maintain ontologies. Ontology learning makes unique contributions to the ontology community by offering efficiency and overcoming the bottleneck in discovering the content of ontologies. In this paper, we proposed a learning-oriented model of ontology development, a framework for ontology learning, and taxonomies of ontology learning approaches and techniques. In view of the impact of domain on ontology learning approaches, we also developed a classification scheme of domains. In order to realize the full potential of ontology learning, we presented current challenges and suggested potential solutions.

Web repositories and recently emerged social networking Web sites feature diversity, large scale, and dynamics, which provides unparalleled valuable resources for ontology learning. As the expansion of the semantic Web, more and more Web content is annotated with semantic information using ontologies, which would ease the task of ontology learning (e.g., [82]). On the other hand, the semantic Web will indirectly benefit from ontology learning by breaking the barrier for wide adoption of the semantic Web.

Ontologies play an instrument role in turning the current Web into a network of knowledge resources and services. Learning ontologies is not an end in itself. In addition to enabling the semantic Web, ontologies may facilitate searching and filtering relevant and trustable information on a subject matter in which a user is interested. As companies and organizations are shifting towards semantic Web services, ontologies become indispensable in expressing request profiles, service profiles, and process models. Ontology learning models may be wrapped as Web services. Service providers are motivated to describe new services and transform legacy services using established ontologies in order to increase the accessibility of their services to consumers. Third parties who are interested in providing add-on services will benefit from ontologies in integrating existing services. An individual who prefers personalized services and content delivery can specify their profiles with popular ontologies such as FOAF (Friend-Of-A-Friend).

References

1. T.R. Gruber, A translation approach to portable ontologies, Knowledge Acquisition 5 (1993) 199–220.
2. N.F. Noy and M.A. Musen, Ontology versioning in an ontology management framework, Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications] 19 (2004) 6–13.
3. A.T. Schreiber, B. Dubbeldam, J. Wielemaker and B. Wielinga, Ontology-based photo annotation, Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications] 16 (2001) 66–74.
4. Z. Duo, L. Juan-Zi and X. Bin, Web Service Annotation Using Ontology Mapping, presented at Service-Oriented System Engineering, 2005. SOSE 2005. IEEE International Workshop (2005).
5. D. Fensel, Ontology-based knowledge management, Computer 35 (2002) 56–59.
6. M. Baziz, M. Boughanem, N. Aussenac-Gilles and C. Chrisment, Semantic cores for representing documents in IR. Proceedings of the 2005 ACM symposium on Applied computing, Santa Fe, New Mexico, ACM Press, 2005, pp. 1011–1017.
7. L. Khan, D. McLeod and E. Hovy, Retrieval effectiveness of an ontology-based model for information selection, The VLDB Journal 13 (2004) 71–85.
8. J. Hendler and D.L. McGuinness, DARPA agent markup language, IEEE Intelligent Systems 15 (2001) 72–73.

9. H. Takeda, K. Iino and T. Nishida, Ontology-supported agent communication, presented at Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments (1995).
10. H. Takeda, K. Iwata, M. Takaai, A. Sawada and T. Nishida, An ontology-based cooperative environment for real-world agents, presented at Second International Conference on Multiagent Systems (1996).
11. M.F. Lopez, A. Gomez-Perez, J.P. Sierra and A.P. Sierra, Building a chemical ontology using methontology and the ontology design environment, Intelligent Systems and Their Applications, IEEE [see also IEEE Intelligent Systems] 14 (1999) 37–46.
12. R. Navigli, P. Velardi and A. Gangemi, Ontology learning and its application to automated terminology translation, Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications] 18 (2003) 22–31.
13. N. Guarino, Formal ontology, conceptual analysis and knowledge representation, International Journal of Human-Computer Studies 43 (1995) 625–640.
14. R. Studer, R. Benjamins and D. Fensel, Knowledge engineering: Principles and methods, Data and Knowledge Engineering 25 (1998) 161–197.
15. D. Brickley and R. Guha, Resource description framework (RDF) schema specification 1.0, vol. 2000: W3C recommendation (2000).
16. D.L. McGuinness and F.v. Harmelen, Web Ontology Language (OWL): Overview (2003).
17. M. Minskey, A framework for representing knowledge, in: P.H. Winston, (ed.) *The Psychology of Computer Vision* (McGraw-Hill, New York, 1975).
18. Y. Ding, Ontology research and development part1 – A review of ontology generation, *Journal of Information Science*, 28 (2002) 123–136.
19. A. Gomez-Perez, Some ideas and examples to evaluate ontologies, presented at Artificial Intelligence for Applications, 1995. Proceedings of 11th Conference on 1995.
20. S. Staab, A. Gomez-Perez, W. Daelemana, M.-L. Reinberger and N.F. Noy, Why evaluate ontology technologies? Because it works!, Intelligent Systems, IEEE 19 (2004) 74–81.
21. Y. Sure, S. Staab and R. Studer, Methodology for development and employment of ontology based knowledge management applications, SIGMOD Rec 31 (2002) 18–23.
22. L. Ding, T. Finin, A. Joshi, R. Pan, R.S. Cost, J. Sachs, V. Doshi, P. Reddivari and Y. Peng, Swoogle: A Search and Metadata Engine for the Semantic Web, presented at Thirteenth ACM Conference on Information and Knowledge Management (CIKM), Washington DC, 2004.
23. A. Maedche and S. Staab, Ontology learning for the Semantic Web, IEEE Intelligent Systems: Special Issue on the Semantic Web 16 (2001) 72–79.
24. M. Uschold and M. Gruninger, Ontologies: principles, methods, and applications, Knowledge Engineering Review 11 (1996) 93–155.
25. D. Faure and C. Nedellec, A corpus-based conceptual clustering method for verb frames and ontology, presented at LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications, 1998.
26. G. Bisson, C. Nedellec and L. Canamero, Designing clustering methods for ontology building – The Mo'K workbench, presented at ECAI Ontology Learning Workshop, Seattle, WA, 2000.
27. P. Buitelaar, D. Olejnik and M. Sintek, A protégé plug-in for ontology extraction from text based on linguistic analysis, presented at the 1st European Semantic Web Symposium (ESWS), Heraklion, Greece, 2004.
28. M. Missikoff, R. Navigli and P. Velardi, Integrated approach to web ontology learning and engineering, IEEE Computer (2002) 60–63.
29. M. Denny, Ontology building: A survey of editing tools, 2002.
30. A. Farquhar, R. Fikes and J. Rice, The Ontolingua server: a tool for collaborative ontology construction, International Journal of Human-Computer Studies 46 (1997) 707–728.
31. M. Missikoff and X.F. Wang, Consys – a group decision-making support system for collaborative ontology building, presented at International Conference on Group Decision and Negotiation, 2001.
32. L. Zhou, Q.E. Booker and D. Zhang, ROD – toward rapid ontology development for underdeveloped domains, presented at 35th Hawaii International Conference on System Sciences, Big Island, Hawaii, 2002.
33. N.F. Noy, M. Sintek, S. Decker, M. Crubezy, R.W. Fergerson and M.A. Musen, Creating semantic web contents with protege-2000, IEEE Intelligent Systems 16 (2001) 60–71.
34. L. Zhou, Q. Booker and D. Zhang, ROD – toward rapid ontology development for underdeveloped domains, presented at 35th Hawaii International Conference on System Sciences (HICSS'35), Hawaii, USA, 2002.
35. S. Soderland, Learning to extract text-based information from the world wide web, presented at Third International Conference on Knowledge Discovery and Data Mining (KDD-97), 1997.
36. J. Allen, *Natural Language Understanding* (The Benjamin / Cummings Publishing Company, Inc., 1995).
37. G. Salton and M. McGill, *Introduction to Modern Information Retrieval* (McGraw-Hill Book Company, 1983).
38. T. Hofmann and J. Puzicha, Statistical Models for Co-occurrence Data, AIM-1625 (1998) 21.
39. B. Roark and E. Charniak, Noun-Phrase Co-occurrence Statistics for Semi-Automatic Semantic Lexicon Construction, presented at COLING-ACL, 1998.
40. E. Riloff and J. Shepherd, A corpus-based bootstrapping algorithm for semi-automated semantic lexicon construction, Journal of Natural Language Engineering 5 (1999) 147–156.
41. P. Jacob and U. Zernik, Acquiring lexical knowledge from text: A case study, presented at 7th National Conference on Artificial Intelligence, 1988.
42. J.-U. Kietz, A. Mädche and R. Volz, A method for semi-automatic ontology acquisition from a corporate intranet, presented at Learning Language in Logic Workshop (LLL-2000), New Brunswick, NJ, 2000.
43. W.B. Dolan, L. Vanderwende and S.D. Richardson, Automatically deriving structured knowledge bases from on-line dictionaries, presented at 1st Conference of the Pacific Association for Computational Linguistics, Vancouver, 1993.
44. A. Shaikevich, Automatic construction of a thesaurus from explanatory dictionaries, Automatic Documentation and Mathematical Linguistics 19 (1985) 76–89.
45. S. Staab, An overview on machine learning for the semantic web, presented at Machine Learning for the Semantic Web, Schloss Dagstuhl, Waden, Germany, 2005.
46. M.E. Califf and R.J. Mooney, Bottom-up relational learning of pattern matching rules for information extraction, Machine Learning Research 4 (2004) 177–210.
47. M. Sanderson and B. Croft, Deriving concept hierarchies from text pages, presented at 22nd ACM SIGIR Conference, 1999.
48. K. Church and P. Hanks, Word association norms, mutual information, and lexicography, Computational Linguistics 16 (1989).
49. S.A. Caraballo, Automatic construction of a hypernym-labled noun hierarchy, presented at 37th Annual Meeting of the Association for Computational Linguistics, 1999.
50. A. Maedche and S. Staab, Mining ontologies from text, presented at 12th International Workshop on Knowledge Engineering and Knowledge Management, French Riviera, 2000.

51. F. Xu, D. Kurz, J. Piskorski and S. Schmeier, Term extraction and mining of term relations from unrestricted texts in the financial domain, presented at Business Information Systems, Poznan, Poland, 2002.
52. D. Lin, Automatic retrieval and clustering of similar words, presented at COLING-ACL'98, Montreal, Canada, 1998.
53. K. Morik, S. Wrobel, J.-U. Kietz and W. Emde, *Knowledge Acquisition and Machine Learning: Theory, Methods, and Applications*. (Academic Press, London, 1993).
54. C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. (MIT Press, Cambridge, MA, 1999).
55. L. Marquez, Machine learning and natural language processing, 2000.
56. G. Ruge, Experiments on linguistically based term associations, Information Processing & Management 28 (1992) 317–332.
57. H. Schutze and J. Pederssn, A cooccurrence-based thesaurus and two applications to information retrieval, presented at RIAO'94, New York, 1994.
58. C. Aone and S.W. Bennett, Evaluating automated and manual acquisition of anaphora resolution strategies, presented at Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, Cambridge, MA, 1995.
59. C. Barbu and R. Mitkov, Evaluation tool for rule-based anaphora resolution methods, presented at Meeting of the Association for Computational Linguistics, Toulouse, 2001.
60. S.K.M. Wong, Y.J. Cai and Y.Y. Yao, Computation of term associations by a neural network, presented at SIGIR'93, Pittsburgh, PA, 1993.
61. K.W. Church and W.A. Gale, A comparison of the enhanced Good-Turing and deleted estimation methods for estimating probabilities of English bigrams, Computer Speech and Language 5 (1991) 19–54.
62. F. Jelinek and R. Mercer, Probability distribution estimation from sparse data, IBM Technical Disclosure Bulletin 28 (1985) 2591–2594.
63. P. Brown, V.D. Pietra, P. deSouza, J. Lai and R. Mercer, Class-based n-gram models of natural language, Computational Linguistics 18 (1992) 467–479.
64. D. Hindle, Noun Classification from predicate-argument structures, presented at ACL-90, Pittsburg, Pennsylvania, 1990.
65. I. Dagan, S. Marcus and S. Markovitch, Contextual word similarity and estimation from sparse data, presented at ACL'93, Columbus, Ohio, 1993.
66. F.C.N. Pereira, N. Tishby and L. Lee, Distributional clustering of English words, presented at 30th Annual Meeting of the ACL, 1993.
67. E. Glover, D.M. Pennock, S. Lawrence and R. Krovetz, Inferring hierarchical descriptions, presented at 11th International Conference on Information and Knowledge Management (McLean, Virginia, 2002).
68. H. Li and N. Abe, Clustering words with the MDL principle, Journal of Natural Language Processing 4 (1997).
69. A. Doan, P. Domingos and A.Y. Levy, Learning source descriptions for data integration, presented at the International Workshop on The Web and Databases (WebDB), Berlin, 2000.
70. Z. Tari, O.A. Bukhres, J. Stokes and S. Hammoudi, The reengineering of relational databases based on key and data correlations, presented at 7th Conference Database Semantics (DS-7), Chapman & Hall, 1997.
71. M.A. Hearst, Automatic acquisition of hyponyms from large text corpora, presented at Proceedings of the Fourteenth International Conference on Computational Linguistics, Nantes, France, 1992.
72. M.A. Hearst, Automated discovery o WordNet relations, in: C. Fellbaum (ed.) *WordNet: An Electronic Lexical Database* (MIT Press, 1998).
73. E. Riloff and J. Shepherd, A corpus-based approach for building semantic lexicon, presented at the Second Conference on Empirical Methods in Natural Language Processing, 1997.
74. N. Aussenac-Gilles, Supervised text analysis for ontology and terminology engineering, presented at Machine Learning for the Semantic Web, Schloss Dagstuhl, Waden, Germany, 2005.
75. P. Wiemer-Hastings, A.C. Graesser and K. Wiemer-Hastings, Inferring the meaning of verbs from context, presented at 20th Annual Conference of the Cognitive Science Society, Mahwah, NJ, 1998.
76. U. Hahn, K. Schnattinger, Ontology engineering via text understanding, presented at IFIP'98, 15th World Computer Congress, Vienna and Budapest, 1998.
77. U. Hahn and K. Schnattinger, Towards text knowledge engineering, presented at 15th National Conference on Artificial Intelligence (AAAI '98), Madison, Wisconsin, 1998.
78. M.-L. Reinberger, Unsupervised text mining for ontology learning, presented at Machine Learning for the Semantic Web, Schloss Dagstuhl, Waden, Germany, 2005.
79. E. Riloff, Automatically constructing a dictionary for information extraction tasks, presented at 11th National Conference on Artificial Intelligence (AAAI'93), Washington, DC, 1993.
80. J. Jannink and G. Wiederhold, Thesaurus entry extraction from an on-line dictionary, presented at 2nd International Conference on Information Fusion (Fusion-99), Omnipress, Wisconsin, 1999.
81. S. Schlobach, Assertional Mining in Description Logics, in: *Description Logics*, (2000) 237–246.
82. F. Ciravegna, S. Chapman, A. Dingli and Y. Wilks, Learning to harvest information for the semantic web, presented at 1st European Semantic Web Symposium (ESWS), Heraklion, Greece, 2004.
83. D. Faure and C. Nedellec, A corpus-based conceptual clustering method for verb frames and ontology acquisition, presented at LREC-98 Workshop on Adapting Lexical and Corpus Resources to Sublan-guages and Applications, Paris, 1998.
84. H. Assadi, Construction of a regional ontology from text and its use within a documentary system, presented at International conference on Formal Ontology in Information Systems (FOIS'98), Amsterdam, 1998.
85. D. Faure and C. Nedellec, Knowledge acquisition of predicate-argument structures from technical texts using machine learning, presented at Current Developments in Knowledge Acquisition (EKAW-99), 1999.
86. A. Thanopoulos, N. Fakotakis and G. Kokkinakis, Automatic extraction of semantic relations from specialized corpora, presented at COLING'2000, Saarbrucken, 2000.
87. P. Cimiano, A. Hotho and S. Staab, Comparing conceptual, divisive and agglomerative clustering for learning taxonomies from text, presented at European Conference on Artificial Intelligence, Valencia, Spain, 2004.
88. M. Finkelstein-Landau and E. Morin, Extracting semantic relationships between terms: supervised vs. unsupervised methods, presented at International Workshop on Ontological Engineering on the Global Information Infrastructure, Dagstuhl Castle, Germany, 1999.
89. D.Z. Inkpen and G. Hirst, Automatic sense disambiguation of the near-synonyms in a dictionary entry, presented at Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2003), Mexico City, 2003.
90. G. Grefenstette, *Explorations in Automatic Thesaurus Discovery* (Kluwer Academic Publishers, 1994).
91. S. Rydin, Building a hyponymy lexicon with hierarchical structure, presented at SIGLEX Workshop on Unsupervised Lexical Acquisition, ACL'02, Philadelphia, Pennsylvania, 2002.

92. J.H. Gennari, S.W. Tu, T.E. Rothenfluh and M.A. Musen, Mapping domains to methods in support of reuse, International Journal of Human-Computer Studies 41 (1994) 399–424.
93. K. Eilbeck, S.E. Lewis, C.J. Mungall, M. Yandell, L. Stein, R. Durbin and M. Ashburner, The sequence ontology: a tool for the unification of genome annotations, Genome Biology 6 (2005).
94. M. Berland and E. Charniak, Finding parts in very large corpora, presented at ACL-99, 1999.
95. M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery, Learning to extract symbolic knowledge from the World Wide Web, presented at the Fifteenth National Conference on Artificial Intelligence (AAAI98), 1998.
96. H. Cui, M.-Y. Kan and T.-S. Chua, Unsupervised learning of soft patterns for generating definitions from online news, presented at the 13th World Wide Web Conference, 2004.
97. P. Brown, S.D. Pietra, V.D. Pietra and R. Mercer, Word sense disambiguation using statistical methods, presented at Annual Meeting of the ACL, 1991.
98. H. Li and N. Abe, Word clustering and disambiguation based on co-occurrence data, presented at COLING-ACL'98, 1998.
99. H.T. Ng and H.B. Lee, Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based approach, presented at 34th Annual Meeting of the Association for Computational Linguistics (ACL), 1996.
100. T. Pedersen and R. Bruce, Knowledge lean word-sense disambiguation, presented at 15th National Conference on Artificial Intelligence, 1998.
101. H. Schutze, Automatic word sense discrimination, Computational Linguistics 24 (1998) 97–123.
102. M. Stevenson and Y. Wilks, The interaction of knowledge sources in word sense disambiguation, Computational Linguistics 27 (2001) 321–350.
103. J.R. Quinlan, Learning logical definitions from relations, Machine Learning Journal 5 (1990) 239–266.
104. C.J. Crouch and B. Yang, Experiments in automatic statistical thesaurus construction, presented at 15th ACM Annual International SIGIR, Denmark, 1992.
105. N. Ge, J. Hale and E. Charniak, A statistical approach to anaphora resolution, presented at Proceedings of the Sixth Workshop on Very Large Corpora (COLING-ACL '98), Montreal, Canada, 1998.
106. W.M. Soon, D. Chung and Y. Lim, A machine learning approach to coreference resolution of noun phrases, Computational Linguistics 27 (2001) 521–544.
107. B. Swartout, R. Patil, K. Knight and T. Russ, Toward distributed use of large-scale ontologies, presented at AAAI-97 Symposium on Ontological Engineering, 1997.
108. P. Buitelaar, Position paper on machine learning for the semantic web, presented at Machine Learning for the Semantic Web, Schloss Dagstuhl, Waden, Germany, 2005.
109. F. Ciravegna, A. Dingli, D. Guthrie and Y. Wilks, Integrating information to bootstrap information extraction from web sites, presented at IJCAI Workshop on Information Integration on the Web, Acapulco, Mexico, 2003.
110. F. Hakimpour and A. Geppert, Resolving semantic heterogeneity in schema integration: an ontology based approach, 2001.
111. N. Noy and M. Musen, An algorithm for merging and aligning ontologies: automation and tool support, presented at Workshop on Ontology Management, Sixteenth National Conference on Artificial Intelligence (AAAI-99), Orlando, FL, 1999.
112. G. Stumme and A. Maedche, FCA-Merge: a bottom-up approach for merging ontologies, presented at 17th International Joint Conference on Artificial Intelligence (IJCAI'01), San Francisco, 2001.
113. S. Dasmahapatra, Interpretation functions for web semantics via machine learning, presented at Machine Learning for the Semantic Web, Schloss Dagstuhl, Waden, Germany, 2005.
114. T. Dietterich, Machine learning research: Four current directions, AI Magazine 18 (1997) 97–136.
115. N. Guarino, Understanding, building and using ontologies, International Journal of Human-Computer Studies 46 (1997) 293–310.
116. D.B. Lenat, CYC: toward programs with common sense, Communications of ACM 33 (1995) 30–49.
117. A. Condamines and J. Rebeyrolles, CTKB: a corpus based approach for terminological knowledge base, presented at Workshop COGNITERM'98 associated to COLING-98, Montreal, Canada, 1998.
118. W.H.E. Davies and P. Edwards, Distributed learning: an agent-based approach to data-mining, presented at ML95 Workshop on Agents that Learn from Other Agents, Tahoe City, California, 1995.
119. L.-K. Soh, Multiagent distributed ontology learning, presented at the First International Joint Conference on Autonomous Agents and Multi-Agent Systems, Workshop on Ontologies in Agent System, Bologna, Italy, 2002.
120. A.B. Williams and C. Tsatsoulis, An instance-based approach for identifying candidate ontology relations within a multi-agent system, presented at Fourteenth European Conference on Artificial Intelligence, Ontology Learning ECAI-2000 Workshop, Berlin, Germany, 2000.
121. A.B. Williams, Learning to share meaning in a multi-agent system, Journal of Autonomous Agents and Multi-Agent Systems 8 (2004) 165–193.
122. U. Guntzer, G. Juttner, G. Seegmuller and F. Sarre, Automatic thesaurus construction by machine learning from retrieval sessions, Information Processing & Management 25 (1989) 265–273.
123. S. Santini, A. Gupta and R. Jain, Emergent semantics through interaction in image databases, IEEE Transactions on Knowledge and Data Engineering 13 (2001) 337–351.