Journal of Southeast University (English Edition) Vol. 22, No. 3, pp. 353 – 356 Sept. 2006 ISSN 1003—7985

Thesaurus-based approach for building domain ontology with a case study of military aircraft prototype ontology construction

Zhen Zhen[1] Tang Aimin[2] Shen Junyi[1] Lu Shengjun[2]

([1] School of Electric and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)
([2] China Defense Science and Technology Information Center, Beijing 100036, China)

Abstract: To alleviate the amount of work involved in constructing a domain ontology, starting with the base of an existing terminological-rich thesaurus is better than starting from scratch. With a case study of reengineering the *Defense Science and Technology Thesaurus* into a prototype military aircraft ontology, a four-phase thesaurus-based methodology is introduced and investigated, which consists of identifying the application purpose, overall design, designing in detail and evaluation. Designing in detail is the core step, converting the terms and semantic relationships of the thesaurus into an ontology and supplementing richer semantic relationships. The resulting prototype ontology includes 87 concepts and 34 relationships, and can be extended and scaled up to a full-fledged domain ontology in the future. Eight universal genres of relationships of this ontology are preliminarily summarized and analyzed, including equivalent relationships, approximate relationships, generic/abstract relationships, part/whole relationships, cause/effect relationships, entity/location relationships etc., and the normalization of semantic relationships is critical to the merging and reusing of follow-up multiple ontologies.

Key words: thesaurus; domain ontology; construction; methodology; semantic relationship

Nowadays, although ontology construction methodologies have not been standardized, there are numerous frequently quoted approaches which are derived mostly from particular ontology building projects. Hence, they are mostly application-dependent or semi-application-dependent[1], with the exception of METHONTOLOGY[2] and the Uschold and King's method[3].

In this paper, we propose a thesaurus-based domain ontology construction approach, which is inspired by the enterprise approach[4-5], the METHONTOLOGY, waterfall software development model by Royce[6], and especially those research efforts involved in the agricultural ontology service (AOS) project by the food and agriculture organization (FAO) of the United Nations[7-8].

1 Thesaurus-Based Methods of Domain Ontology Construction

1.1 Advantage of thesaurus-based approach

The thesaurus has been introduced and then widely used in full- or semi-automatic index term selection, or as a post-controlling assistant to improve information retrieval[9], since the first thesaurus controlling the vocabulary in an information retrieval system was complied by the Du Pont organization in 1959.

As defined in ISO 2788[10], a thesaurus is "the vocabulary of a controlled indexing language, formally organized so that *a priori* relationships between concepts (for example as "broader" and "narrower") are made explicit."

The thesaurus consists of terms (descriptors and non-descriptors) and relationships between them, using a set of indicators to display and distinguish these relationships. Since the *Thesaurus of Engineering and Scientific Terms* (TEST) was published in 1967, three kinds of semantic relationships, i. e., equivalence, hierarchical, and associative relationships have been used in the thesaurus[11].

The first definition of ontology from the knowledge community, excepting philosophy, according to Corcho et al.[1], was made by Neches and his colleagues, while the most quoted definition was given by Gruber in 1993[12], as "an explicit specification of a conceptualization". Domain ontology, one kind of ontology, provides definitions and relationships of the concepts, and major theories and principles and activities in the domain.

Both the thesaurus and the ontology consist of terms (concepts or classes) and the relationships be-

Received 2006-04-25.

Biographies: Zhen Zhen (1970—), male, graduate, zhenzh@mailst.xjtu.edu.cn; Shen Junyi (1939—), male, professor, jyshen@mail.xjtu.edu.cn.

tween terms (concepts or classes), and, furthermore, the thesaurus for a specific field usually includes a relatively complete set of terms in that field collected and organized by domain experts and knowledge organization professionals, and these terms can be used as concept candidates in a domain ontology. Qualifiers, scope notes, relationships in the thesaurus can also be used as property candidates, instance candidates and relationship candidates in the domain ontology, saving the ontology developers from re-collecting and re-organizing them.

1.2 Outline of our approach

The approach consists of four phases: ① To identify the application purpose of the domain ontology, ② To design it overall, ③ To design it in detail, ④ To evaluate it. The result of one phase will be verified before it becomes the input of next phase. If it is confirmed, then goes to the next phase; otherwise, it goes back to the current phase or even the prior phase. We will further illustrate phase ③ in detail in the next section with a case study of reengineering the *Defense Science and Technology Thesaurus* into a prototype military aircraft ontology.

2 Case Study of Reengineering DSTT

2.1 Background of DSTT

The *Defense Science and Technology Thesaurus* (DSTT) is a multilingual thesaurus designed to cover concepts and terminology in defense science and technology related domains (e.g. aerospace). This thesaurus was developed by the China Defense Science and Technology Information Center (CDSTIC) and some other defense industry information centers in China in the early 1990s, containing approximately 63×10^3 descriptors and 32×10^3 non-descriptors.

2.2 Phase one and two

In these two phases, the application purpose, the developing tools and the representation language should be determined, as well as the bounds, constraints, and principles of the ontology.

In our case, the main purpose of the prototype military aircraft domain ontology is a testbed for our method and should include only the essential elements. Although it is not a full-fledged domain ontology, it should be extensible and scalable to be embedded into a "real" application. We select KAON as the developing tool and RDFS as the representation language. KAON is an environment originally developed by the University of Karlsruhe, Germany, and now shifted to the open-sourced community and freely available.

2.3 Designing domain ontology in detail

2.3.1 Converting terms and their hierarchical relationships

We adopt the middle-out method and select the most important concept "military aircraft" as the first concept and then the less important concepts such as "pilot", "aircraft engine", "airborne weapon" etc. from DSTT, also the related hierarchical relationships of these terms are selected, for example, the sub-concepts of "military aircraft" such as "fighter aircraft", "bomber aircraft", and "reconnaissance aircraft", etc.

We also notice that only some of the terms of DSTT can be converted to concepts or sub-concepts of the ontology. The others may be converted to the property of concepts. The terms that represent concrete objects are more likely to be converted to concepts of the ontology, and the terms that represent the character, phenomenon, status or procedure of objects are more likely to be converted to the properties or relationships of concepts of the ontology.

2.3.2 Converting equivalent relationships

All the synonyms are put into their corresponding preferred node to avoid loss of information, without any relationships. In our case, we put the terms under UF (used for) into the "synonym" column of the Lexicon of the corresponding node. For example, "interceptor aircraft" and "air combat fighter" are under UF of "fighter aircraft" in DSTT, so we copy them into the "synonym" column of the Lexicon of the concept of "fighter aircraft" in the target ontology.

2.3.3 Converting associate relationship

We used to think the associate relationship in DSTT would help us to designate the non-hierarchical relationships between concepts of the ontology, but unfortunately we were wrong. The main cause of this problem is that so many related terms (RT) are missing under associated relationship that we cannot count on the Thesaurus to list all the related terms, and furthermore, DSTT cannot help us to select the relationship tag either. For example, there is only one term "military aviation" associated with the term "military aircraft" in DSTT.

2.3.4 Supplementing other properties

We supplement other properties from many sources, for example, "ceiling", "crashworthiness" etc. under "0401 aeronautical general" from the scope and subject category lists of DSST, and "endurance", "wing span" in the hierarchical list of the Thesaurus, and "aircraft height", "aircraft length" from some other handbooks. These terms are added as the properties of "military aircraft".

2.3.5 Adding other relationships between concepts

In KAON, non-hierarchical and non-equivalent relationships are regarded as special properties connecting two concepts, and we call these special properties as relationship properties and the normal properties as value properties. In KAON all properties are displayed as a vector with not only a value but also a direction. The relationships between concepts depend so much on domain knowledge that we mainly refer to handbooks, dictionaries, and domain experts to determine them.

3 Analyzing Semantic Relationships

Most of the relationships between concepts in our ontology can be classified into the following eight genres[7, 13]. We use a notion of *A* 〈*X*〉 *B* representing that concept *A* has a relationship of *X* with concept *B*.

① Equivalent relationship (*A* 〈equal to〉 *B*, then *B* 〈equal to〉 *A*)

This relationship is applicable to a constellation where concept *A* has the same meaning as concept *B*. For example, fighter aircraft 〈equal to〉 interceptor aircraft.

② Approximate relationship (*A* 〈similar to〉 *B*, then *B* 〈similar to〉 *A*)

This relationship is applicable to a constellation where concept *A* has a similar meaning to concept *B*. For example, military engineering 〈similar to〉 defense engineering.

③ Generic/abstract relationship (*A* 〈includes specific〉 *B*, then *B* 〈is a〉 *A*)

This relationship is applicable to a constellation where concept *A* is a broader concept of concept *B*. For example, military aircraft 〈includes specific〉 fighter aircraft, fighter aircraft 〈is a〉 military aircraft.

④ Part/whole relationship (*A* 〈has component〉 *B*, then *B* 〈is component of〉 *A*)

This relationship is applicable to a constellation where concept *A* has a component of concept *B*, which retains its identity as an object even when built into the whole. For example, military aircraft 〈has component〉 aircraft engine, aircraft engine 〈is component of〉 military aircraft.

⑤ Cause/effect relationship (*A* 〈causes〉 *B*, then *B* 〈caused by〉 *A*)

This relationship is applicable in a case where concept *A* causes concept *B* to happen. For example, bird strike 〈causes〉 crash, crash 〈caused by〉 bird strike.

⑥ Entity/location relationship (*A* 〈located in〉 *B*, then *B* 〈holds〉 *A*)

This relationship is applicable in a case where concept *A* is located in concept *B*. For example, pilot 〈located in〉 pilot seat, pilot seat 〈holds〉 pilot.

⑦ Agent/object relationship (*A* 〈operates〉 *B*, then *B* 〈operated by〉 *A*)

This relationship is applicable in a case where concept *A* has at least one kind of operation on/to concept *B*, and the operation type usually needs further qualification. For example, pilot 〈operates type = "manoeuvre"〉 military aircraft, military aircraft 〈operated by type = "manoeuvre"〉 pilot.

⑧ Execute/method relationship (*A* 〈executes〉 *B*, then *B* 〈executed by〉*A*)

This relationship is applicable in a case where concept *A* executes concept *B*, and the execution method usually needs further qualification. For example, military aircraft 〈executes method = "manual"〉 flight, flight 〈executed by method = "manual"〉 military aircraft.

This preliminary analysis shows that our ontology contains many kinds of relationships that can be normalized for future reuse, and that a more complete inventory of relationship types is needed when multiple ontologies are to be merged and reused.

4 Conclusion

With the increase of concepts included in the domain ontology, labor involved in ontology construction will increase dramatically. Besides the necessity that the scope of the domain ontology and the construction team members should be determined carefully, we suggest that the construction project ontology should start at several smaller domains, and then develop a larger domain ontology incrementally via merging, integrating and re-using these small pieces of ontologies.

We still leave some interesting questions to be answered such as: How can the relationships in a thesaurus be translated into a semantic-rich domain ontology with better precision and accuracy? And, will we be able to perform the transformation task full-or semi-automatically, and how?

References

[1] Corcho O, Fernández-López M, Gómez-Pérez A. Methodologies, tools and languages for building ontologies: where is their meeting point[J]. *Data & Knowledge Engineering*, 2003, **46**(1): 41 − 64.

[2] Fernández-López M, Gómez-Pérez A, Pazos-Sierra A, et al. Building a chemical ontology using METHONTOLOGY and the ontology design environment [J]. *IEEE Intelligent Systems & Their Applications*, 1999, **14**(1): 37 - 46.
[3] Uschold M, King M. Towards a methodology for building ontologies [EB/OL]. (1995-07-31) [2006-05-07]. http://www.aiai.ed.ac.uk/project/ftp/documents/1995/95-ont-ijcai95-ont-method.ps.
[4] Uschold M. Building ontologies: towards a unified methodology[EB/OL]. (1996-09-01) [2006-05-07]. http://www.aiai.ed.ac.uk/project/pub/documents/1996/96-es96-unified-method.ps.
[5] Uschold M, Gruninger M, Ontologies: principles methods and applications [J]. *The Knowledge Engineering Review*, 1996, **11**(2): 93 - 155.
[6] Royce W W. Managing the development of large software systems: concepts and techniques [A]. In: *Proc of the 9th international conference on Software Engineering* [C]. Los Alamitos: IEEE computer Society Press, 1987. 328 - 338.
[7] Soergel D, Lauser B, Liang A, et al. Reengineering thesauri for new applications: the AGROVOC example [EB/OL]. (2004-03-17) [2006-05-07]. http://jodi.tamu.edu/Articles/v04/i04/Soergel/.
[8] Lauser B, Wildemann T, Poulos A, et al. A comprehensive framework for building multilingual domain ontologies: creating a prototype biosecurity ontology [A]. In: *Proc of International Conference on Dublin Core and Metadata for e-Communities* [C]. Florence: Firenze University Press, 2002. 113 - 123.
[9] Li Jing, Qian Ping. Differences and relationships between thesaurus and ontology [J]. *Journal of Library Science in China*, 2004, **30**(1): 36 - 39. (in Chinese)
[10] ISO TC46/SC9. ISO 2788—1986 guideline for the establishment and development of monolingual thesauri [S]. Geneva: International Organization for Standardization, 1986.
[11] Aitchison J, Clarke S D. The thesaurus: a historical viewpoint, with a look to the future [J]. *Cataloging & Classification Quarterly*, 2004, **37**(3, 4): 5 - 21.
[12] Gruber T R. A translation approach to portable ontology specifications [J]. *Knowledge Acquisition*, 1993, **5**(2): 199 - 220.
[13] Schmitz-Esser W. Thesaurus and beyond: an advanced formula for linguistic engineering and information retrieval [J]. *Knowledge Organization*, 1999, **26**(1): 10 - 22.

基于叙词表构建军用飞机领域本体原型

真　溱[1]　唐爱民[2]　沈钧毅[1]　卢胜军[2]

([1] 西安交通大学电子与信息工程学院,西安 710049)
([2] 中国国防科技信息中心,北京 100036)

摘要:为了减少构建领域本体的工作量,利用现有术语丰富的叙词表作为构建工作的起点是一种较好的方法.以《国防科技叙词表》的有关内容转换生成军用飞机领域本体原型为实例,详细叙述了一种基于叙词表构建领域本体的方法.该方法包含4个阶段,即确定用途、总体设计、详细设计和评估.详细设计是核心,该部分将叙词表中出现的词汇、语义关系转换到本体中,并补充了更丰富的语义关系.生成的领域本体原型具有87个概念、34个关系,具备较好的扩展能力.对本体原形中出现的8种通用关系类型也进行了初步地归纳和分析,包括等同关系、近似关系、具体与抽象关系、部分与整体关系、因果关系、实体与位置关系等.语义关系规范化将是今后实现多个本体合并和复用的一个关键.

关键词:叙词表;领域本体;构建;方法论;语义关系
中图分类号:TP311