通过本体演化创建语义增强的云服务环境

曹钟元

大连海事大学 计算机科学与技术学院 计算机技术 1120161172

摘 要 目前，Web资源的可用性已经大大增加，使得无论用户在给定时刻需要什么都可以在互联网上找到。这些资源不再限于数据项，而是通过某种服务架构模型传递的功能在因特网上提供。在过去几年中，云计算已经成为在互联网上提供服务的最流行的计算模型之一。然而，随着可用云服务的数量增加，出现了服务发现和选择的问题。经验表明，语义技术可以为增强和更精确的搜索过程提供基础。在本文中，我们提出一个平台，利用语义技术和技术，以方便发现满足用户的需求的云资源。我们提出了一种架构，其将语义注释技术，本体进化，词语提取和索引资源以语义地注释云服务，以及语义搜索引擎，利用云资源的语义描述，从基于关键字的搜索中找到它们。在ICT领域对该工具的全面评估产生了非常有希望的结果，并在本文中提出。

关键字 语义注释; 本体演化; 云计算; 信息提取; 知识管理; 本体; 语义网

**Creating a semantically-enhanced cloud services environment through ontology evolution**

Zhongyuan Cao

Department of computer, Dalian Maritime University, City Dalian

**Abstract** Currently, the availability of Web resources has grown enormously to the point that whatever a user needs at a given moment can potentially be found on the Internet. These resources are not limited to data items anymore, functionality delivered through some sort of service architectural model is also offered on the Internet. In the last few years, cloud computing has emerged as one of the most popular computing models to provide services over the Internet. However, as the number of available cloud services increases, the problem of service discovery and selection arises. Experience indicates that semantic technologies can provide the basis for enhanced and more precise search processes. In this paper, we present a platform that makes use of semantic technologies and techniques to facilitate the discovery of cloud resources meeting the users’ needs. We propose an architecture that puts together semantic annotation techniques, ontology evolution, term extraction and indexing resources to semantically annotate cloud services, and a semantic search engine that leverages the semantic description of the cloud resources to find them from keyword-based searches. A comprehensive evaluation of the tool in the ICT domain has produced very promising results and is also presented in this article.

**Key words** Semantic annotation; Ontology evolution; Cloud computing; Information extraction; Knowledge management; Ontology; Semantic web

**1.绪论**

专家将云计算定义为“用于实现对可配置计算资源（例如，网络，服务器，存储，应用和服务）的共享池的无处不在的、方便的、按需的网络访问的模型，其可以被快速供应和释放，占用极小的资源或服务提供者需要进行的交互行为”[1]。此外，云计算可以被视为允许按次收费的模型在因特网上提供计算服务的技术范例转移。云的这个后来的特征使其对当前社会经济环境下的公司特别有吸引力。因此，根据最着名的IT分析公司Gartner公司的研究，云计算是IT支出增长的主要来源。国家标准与技术研究所（NIST）成立了一个集团，致力于促进政府和行业内云技术的有效和安全使用。他们确定了三个主要的服务模型，即软件即服务（SaaS），平台即服务（PaaS）和基础设施即服务（IaaS）[1]。前者是模型，代表终端用户实现其目标的实际应用。 PaaS是表示可以构建应用程序的框架和公共函数的模型。后者指的是提供处理，存储，网络和其他基本计算资源的服务，消费者可以在这些资源上部署任意软件。本文中当使用术语“云服务”时，指的是上述三种类别（SaaS，PaaS和IaaS）中的任何一种服务。随着云计算模型的增长势头，越来越多的云服务变得可用。虽然服务消费者可以明显地从这种广泛的服务提供中受益，但是它们被手动服务发现的耗时的任务显着地妨碍。事实上，可用服务数量的指数增长使得用户更难以找到满足他们的需求的服务。因此，云计算社区的一个主要关注点是服务搜索和检索[2][3]。在这些情况下，作者认为在其他环境中应用语义技术以增强搜索过程（例如，语义Web服务[4]）的成功可以无缝地应用在云上下文中。语义网络技术[5]目前正在实现一定程度的成熟。语义Web被设想为向在Web上发布的数据添加语义（建立数据的含义），使得机器能够以人类类似的方式处理这些数据。语义技术应用于旅游业[6][7]、财政[8][9]、电子学习[10]、休闲[11][12]和生物科学[13][14]。本体是语义Web的最重要的技术。一个本体可以被定义为“共享概念化的正式和显式规范”[15]。本体提供正式的、结构化的知识表示，具有可重复使用和可共享的优点。本体中的知识主要使用五种组件形式化：类，关系，属性，公理和实例。本体和语义技术的使用可以克服传统搜索工具的局限性[16][17]，并且它们也被证明在例如Web浏览[18][19]、服务发现[4]、建议[20]和信息管理[21]中的价值。

由于手动寻找合适的服务是耗时的，因此向云环境中添加语义有助于自动化搜索过程[22]。这种方法需要生成所有可用云服务的语义描述。如果手动执行，这将是一个非常耗时的任务。相反，可以将自然语言处理（NLP）技术应用于描述用于向其添加语义注释的服务的文档。计算机然后可以通过利用可用的语义信息来帮助用户发现适当的服务。与此相一致，在本文中，我们提出一个基于本体进化的框架，使云服务可用的语义注释。然后利用附加到云服务的注释，以便在云服务空间上执行语义增强的搜索。最终目标是改进与满足用户需求的云服务的搜索和检索相关的过程。所提出的架构将语义注释技术、本体演化、术语提取和索引资源以语义注释云服务以及语义搜索引擎、利用云资源的语义描述、基于关键字搜索这些技术融合。这项工作的主要贡献是设计和开发用于云服务注释和检索的语义平台。该平台旨在减轻手动发现合适的云服务的负担。为此，提出了以多种方式推进现有技术的语义注释机制。我们的方法提供对多个本体的支持，可以处理非结构化文档，支持文档和本体演化，并且是全自动的。总而言之，我们的提案和其他相关作品之间的主要区别是可以从几种数字格式中提供的自然语言描述中自动注释云服务，从而有助于提高服务搜索过程的精确度。

2.平台构架

所提出的方法的架构如图1所示。系统由五个主要模块组成：（1）语义表示和注释模块，（2）语义索引模块，（3）术语提取模块，（4）本体演化模块和（5）语义搜索引擎。简而言之，系统包括两个单独的阶段：语义云服务存储库的创建和云服务发现。第一阶段工作首先，在云中的服务的自然语言描述被语义地表示和注释。然后，从这些注释中，使用经典向量空间模型创建语义索引。同时，项提取器获得出现在文本中不存在于域本体中的术语。最后，本体演化模块检查先前收集的最重要的新词是否可以添加到域本体。另一方面，对于云服务发现阶段，已经设计了允许从基于关键字的搜索检索匹配服务的语义搜索引擎。接下来我们将详细描述这些部件。

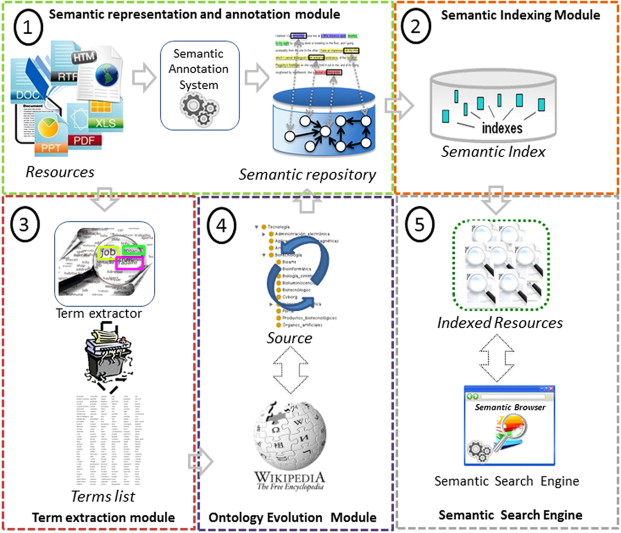


图1：系统功能架构

2.1语义表示和注释模块

GoodRelations [23]是一个OWL本体，用于描述在Web上提供的产品和服务。它可以被视为产品、价格、商店和公司数据的标准化词汇，它现在完全与HTML5微数据规范兼容，并且可以用作schema.org词汇表的电子商务扩展。这个本体在这个工作中用于语义地表示云服务的部分。特别是，它有助于描述公司、价格和其他有关产品或服务的相关信息。当由系统收集关于新云服务的信息时，创建GoodRelations本体的类“ProductOrService”的个体，包括关于该服务的上下文信息。此外，使用引用类的“描述”数据类型属性将服务的自然语言描述添加到本体。一旦个体被引入到本体中，该模块检索服务相关的自然语言描述和领域本体。然后，通过使用一组NLP工具，它根据域本体和Wordnet产生服务描述的语义注释。该模块基于[24]中提出的方法，由两个主要阶段组成：NLP阶段和语义注释阶段。NLP阶段的主要目的是提取每个句子的形态结构。为此，已经开发了一组NLP软件工具，包括句子检测组件，分词器，一组POS标签器，一组缩略语和一组句法分析器。此外，GATE框架2已经被采用。 GATE是用于开发和部署处理人类语言的软件组件的基础设施。 GATE通过三种方式帮助科学家和开发人员：（i）通过指定语言处理软件的架构或组织结构; （ii）通过提供实现架构并且可以用于在不同应用中嵌入语言处理能力的框架或类库; （iii）通过提供构建在由用于开发组件的方便的图形工具组成的框架之上的开发环境。作为该第一阶段的结果，获得表示文本的句法结构的一组注释。在该方法的第二阶段期间，通过遵循下面描述的过程，云服务自然语言描述被注释有域本体的类和实例。首先，最重要的语言表达通过基于文本的句法结构的统计方法来识别。然后，对于每个语言表达式，系统试图确定这样的表达式是否是领域本体的类的个体。

2.2语义索引模块

在此模块中，系统从上一个模块检索所有注释的知识，并尝试使用此知识创建完全填充的注释。 这一步是基于[25]中提出的工作。每个文档的每个注释存储在数据库中并且具有分配的权重，其反映本体实体对文档含义的相关性。通过使用TF-IDF算法[26]计算权重，其使用以下等式。

= ∗ (1)

其中是本体实体i在文档d中出现的次数，是文档d中标识的所有本体实体的出现次数的总和。是所有文档的集合，是用i注释的所有文档的数目。

在这种情况下，云服务描述的是要分析的文档。对于每个描述，基于在[25]中呈现的经典向量空间模型的适应性来计算索引。每个服务被表示为向量，其中每个维度对应于域本体的单独的本体论概念。每个本体概念维度的值计算如下。

where = (2)

其中dist（i，j）是领域本体中概念i和概念j之间的语义距离。该距离通过使用领域本体中的概念的分类学关系来计算。因此一个概念与其本身之间的距离为0，一个概念与其分类父或子之间的距离为1，以此类推。

2.3术语提取模块

通过本单元，识别文本文档中最重要的术语。 在本阶段收集的术语稍后由本体演化模块使用，以便保持领域本体是最新的。假设存在多字和单字项。通过考虑这个假设，已经实现了两种不同的方法：NC-值算法[27]，其允许获得表示概念的多字项候选和TF-IDF已经被用于获得由一个词形成的词语。在文本中出现的不是域本体的一部分的最重要的术语的列表被作为本体演进模块的输入。

2.4 本体演化模块

这个模块的主要目的是使用维基百科中的信息来维护和发展本体。术语提取模块收集的术语列表用于丰富，增强和增加由域本体表示的知识。维基百科是一个自由的百科全书，其中数千个概念根据维基百科的类别分类在分类中。这种方法的主要思想是通过利用维基百科的结构，即将文章分配到类别，子类别关系和跨语言链接等效文章的跨语言关系，来保持领域本体是最新的。维基百科结构是一个重要和必要的条件，因为它允许建立定义语义概念的术语组。DBPedia是一个旨在从作为Wikipedia项目的一部分创建的信息中提取结构化内容的项目。然而，在维基百科中存在的结构和组织在DBPedia中还不能用于所有语言，因为DBPedia专注于表示文章的信息，而不是维基百科类别形成的分类。

在维基百科中查找由术语提取器模块产生的列表中的每个相关术语（即，当前不存在于本体中的术语）。如果发现与该术语匹配的维基百科文章或类别，则创建包含英语和西班牙语中的所有术语的同义词的新概念。该算法已经通过使用OWL API 2框架和Java维基百科API（Bliki引擎3）来实现，其是用于将维基百科wikitext符号转换为HTML的解析器库。该模块设计为允许使用超文本传输​​协议（HTTP）访问在线wikitext。因此，提取的内容总是最新的，因为它是从维基百科的在线版本收集的。此外，如果需要，工具允许收集过时的wikitext内容。

2.5语义搜索引擎

该模块负责通过基于关键字的查询在云中查找服务。该过程利用了系统先前收集的语义内容和注释。首先，用户引入一系列关键字，并且系统识别域本体中的哪些概念由它们引用。如在前面的部分中已经解释的，每个服务被表示为向量，其中每个维度对应于域本体的单独概念。然后，语义搜索引擎计算查询’q’和每个服务's'之间的相似度值。为了做到这一点，使用余弦相似性：

== (3)

然后通过使用相似性函数来定义与查询中引用的主题相关的最相关的云服务的排名。向量空间模型s通过使用等式（2），并且第二向量q是根据从搜索引擎查询提取的概念创建的。符号是分离两个向量的角度，并且表示查询和每个特定云服务之间的相似等级。

3. 总结

云计算是一种允许通过互联网提供计算服务的技术范例。这种新的服务模型与以前的，众所周知的分布式计算举措密切相关，例如Web服务和网格计算。在当前的社会经济气候下，云计算的可负担性已经提高了其在当今的创新中的普及。在这些情况下，越来越多的云服务变得可用。云计算产业的增长导致从云交付的服务数量呈指数增长。作为这种增加的直接结果，人工找到最佳地满足个人或组织的需要的服务所需的时间和精力被显着增强。云服务的语义注释和检索是一项具有挑战性的任务，它解决了找到具有适当功能的正确服务的问题。

在本文中，已经提出了用于云服务注释和从其描述中检索的语义平台。这里提出的系统自动地从不同的云服务注释他们的自然语言描述，这可以在许多文档格式，如XML，HTML或PDF。此外，所提出的平台已经考虑到多方面环境（具有OWL 2本体）来实现，以能够应对若干领域。此外，它支持源文档的演进，从而保持自然语言描述和注释之间的一致性，其使用基于语义的Web的模型存储。此外，连续更新本体以令人满意地反映域中的变化是至关重要的。本体演化是关于本体适应变化需求的及时适应的研究领域。我们的框架包含一个模块，通过利用维基百科的内容，组织和结构解决本体进化。全局最终平台将本体演化方法与基于语义的信息检索系统相结合。该集成方法允许自动保持本体是最新的，使得检索工具在其运行期间不断精确。

重要的是要考虑到，尽管用于构建平台的技术是众所周知的，就我们而言，它们在云计算领域中的联合应用还没有被探索。实际上，这种特定的组合帮助我们实现了我们的初始目标，开发语义云服务存储库，以及实现精确和自动机制来发现满足一些约束的云服务。我们的建议相对于其他相关工作的主要优势在于集合了以前不支持的一些解决方案和功能。此外，这里提出的系统解决了更一般的问题，即从非结构化文档生成语义注释。

几个问题仍然未来工作。到目前为止，已经通过探索其自然语言描述来分析服务。通过使用可以描述这些服务的本体，如[22]所示，使用关于它们的功能的语义信息也是有益的。此外，我们正在努力升级此系统，并将其转换为推荐系统，用户可以在其中设置其首选项，系统将只返回特定域中的相关服务。此外，我们计划研究提供搜索服务的可能性，还包括意见挖掘引擎，如[28]中提出的，允许获得服务的感性分类，以提供有关其非功能属性。最后，语义搜索模块限于发出基于关键字的查询。语义搜索者的当前方法使用SPARQL，RDQL和OWLQL，但是它们太复杂，不能被非专家用户使用。对于未来的工作，计划在平台中包括基于我们研究小组以前的作品来查询系统的自然语言接口[29]。

参考文献

[1] P. Mell, T. Grance, The NIST definition of cloud computing, NIST Special Publications (800 Series), SP 800-145, 2011. Available at: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[2] J. Kang, K. Mong Sim, Cloudle: a multi-criteria cloud service search engine, in: Proc. Services Computing Conference, APSCC, 2010, pp. 339–346.

[3] M. Wang, J. Zhou, S. Jing, Cloud manufacturing: needs, concept and architecture, in: Proc. Computer Supported Cooperative Work in Design, CSCWD, 2012, pp. 321–327.

[4] F. García-Sánchez, R. Valencia-Garcia, J.T. Fernández-Breis An ontology, intelligent agent-based framework for the provision of semantic Web services Expert Systems with Applications, 36 (2) (2009), pp. 3167–3187

[5] N. Shadbolt, T. Berners-Lee, W. Hall The semantic Web revisited IEEE Intelligent Systems, 21 (3) (2006), pp. 96–101

[6] J.M. Ruiz-Martínez, J.A. Miñarro-Giménez, D. Castellanos-Nieves, F. García-Sánchez, R. Valencia-García

Ontology population: an application for the e-Tourism domain

International Journal of Innovative Computing, Information and Control, 7 (11) (2011), pp. 6115–6134

[7] A. García-Crespo, J. Chamizo, I. Rivera, M. Mencke, R. Colomo-Palacios, J.M. Gómez

SPETA: social pervasive e-Tourism advisor

Telematics and Informatics, 26 (3) (2009), pp. 306–315

[8] A. Esteban-Gil, F. García-Sánchez, R. Valencia-Garcia, J.T. Fernandez-Breis

SocialBROKER: a collaborative social space for gathering semantically-enhanced financial information

Expert Systems with Applications, 39 (10) (2012), pp. 9715–9722

[9] A. Rodríguez-González, R. Colomo-Palacios, F. Guildrís-Iglesias, J.M. Gómez-Berbís, A. García-Crespo

FAST: fundamental analysis support for financial statements, using semantics for trading recommendations

Information Systems Frontiers, 14 (4) (2012), pp. 999–1017

[10] X. Luo, X. Wei, J. Zhang

Guided game-based learning using fuzzy cognitive maps

IEEE Transactions on Learning Technologies, 3 (4) (2010), pp. 344–357

[11] W. Carrer-Neto, M.L. Hernández-Alcaraz, R. Valencia-García, F. García-Sánchez

Social knowledge-based recommender system, application to the movies domain

Expert Systems with Applications, 39 (12) (2012), pp. 10990–11000

[12] A. García-Crespo, R. Colomo-Palacios, J.M. Gómez-Berbís, F. García-Sánchez

SOLAR: social link advanced recommendation system

Future Generation Computer Systems, 26 (3) (2010), pp. 374–380

[13] J.M. Ruiz-Martínez, R. Valencia-García, J.T. Fernández-Breis, F. García-Sánchez, R. Martínez-Béjar

Ontology learning from biomedical natural language documents using UMLS

Expert Systems with Applications, 38 (10) (2011), pp. 12365–12378

[14] A. Rodríguez-González, J.E. Labra-Gayo, R. Colomo-Palacios, M.A. Mayer, J.M. Gómez-Berbís, A. García-Crespo

SeDeLo: using semantics and description logics to support aided clinical diagnosis

Journal of Medical Systems, 36 (4) (2012), pp. 2471–2481

[15]. Studer, V.R. Benjamins, D. Fensel

Knowledge engineering: principles and methods

Data & Knowledge Engineering, 25 (1998), pp. 161–197

[16] Z. Xu, X. Luo, J. Yu, W. Xu

Mining Web search engines for query suggestion

Concurrency and Computation: Practice and Experience, 23 (10) (2011), pp. 1101–1113

[17] E. Lupiani-Ruiz, I. Garcia-Manotas, R. Valencia-García, F. García-Sánchez, D. Castellanos-Nieves, J.T. Fernández-Breis, J.B. Camón-Herrero

Financial news semantic search engine

Expert Systems with Applications, 38 (12) (2011), pp. 15565–15572

[18] X. Luo, Z. Xu, J. Yu, X. Chen

Building association link network for semantic link on Web resources

IEEE Transactions on Automation Science and Engineering, 8 (3) (2011), pp. 482–494

[19] Z. Xu, X. Luo, J. Yu, W. Xu

Measuring semantic similarity between words by removing noise and redundancy in Web snippets

Concurrency and Computation: Practice and Experience, 23 (18) (2011), pp. 2496–2510

[20] A. García-Crespo, J.L. López-Cuadrado, R. Colomo-Palacios, I. Gonzalo-Carrasco, B. Ruiz-Mezcua

Sem-Fit: a semantic based expert system to provide recommendations in the tourism domain

Expert Systems with Applications, 38 (10) (2011), pp. 13310–13319

[21] R. Colomo-Palacios, A. Garcia-Crespo, P. Soto-Acosta, M. Ruano-Mayoral, D. Jiménez-López

A case analysis of semantic technologies for R&D intermediation information management

International Journal of Information Management, 30 (5) (2010), pp. 465–469

[22] G. Ortegón-Cortázar, J.J. Samper-Zapater, F. García-Sanchez, Adding semantics to cloud computing to enhance service discovery and access, in: Proceedings of the 6th Euro American Conference on Telematics and Information Systems, EATIS 2012, Valencia, Spain, 2012.

[23] M. Hepp

GoodRelations: an ontology for describing products and services offers on the Web

Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management, EKAW2008, Springer (2008), pp. 332–347

[24] M. Vargas-Vera, E. Motta, J. Domingue, M. Lanzoni, A. Stutt, F. Ciravegna

MnM: ontology driven semi-automatic and automatic support for semantic markup ,in: A. Gómez-Pérez, V.R. Benjamins (Eds.), EKAW 2002, LNCS (LNAI), 2473 (379) (2002), pp. 379–391

[25] P. Castells, M. Fernández, D. Vallet

An adaptation of the vector-space model for ontology-based information retrieval

IEEE Transactions on Knowledge and Data Engineering, 19 (2) (2007), pp. 261–272

[26] G. Salton, M.J. McGill

Introduction to Modern Information Retrieval

0070544840McGraw-Hill (1983)

[27] A. Barrón-Cedeño, G. Sierra, P. Drouin, S. Ananiadou

An improved automatic term recognition method for Spanish

Lecture Notes in Computer Science, vol. 5449 (2009), pp. 126–136

[28] I. Peñalver-Martínez, R. Valencia-García, F. García-Sánchez, Ontology-guided approach for feature-based opinion mining, in: NLDB, Alicante, Spain, 2011, pp. 193–200.

[29] R. Valencia-García, F. García-Sánchez, D. Castellanos-Nieves, J.T. Fernández-Breis

OWLPath: an OWL ontology-guided query editor

IEEE Transactions on Systems, Man, and Cybernetics Part A, 41 (1) (2011), pp. 121–136