

顶级专家Frank van Harmelen 揭秘语义网

◎ 译/徐涵

所谓“语义”就是文本的含义。语义需要理解文本的意思和结构，而与显示方式无关。简单地说，语义网（Semantic Web）是一种能理解人类语言的智能网络，它不但能够理解人类的语言，而且还可以使人与电脑之间的交流变得像人与人之间交流一样轻松。但目前大多数人对语义网的认识都还存在误区，因此我们特编辑此文为语义网揭秘。

记者：常有人说，语义网所要解决的问题，就是30年前人工智能（Artificial Intelligence）里的知识表示（Knowledge Representation）及归纳逻辑（Inductive Logics）所要解决的问题。KL-ONE、Cyc以及Minsky的框架（Frames）和Sowa的概念图（Conceptual Graphs）等，都属于过去这些工作留下的产物。但是它们都已经失败了。那么语义网以及语义网在本体（Ontology）^①和推理（Reasoning）方面的关注，跟这些失败的努力有什么不同呢？

Frank：其实，大家对语义网存在一种误解，即认为语义网是“重复人工智能的工作”。虽然语义网和人工智能（AI）所用的工具有一些相同（比如本体、推理、逻辑等），但它们的目标是完全不同的。实际上，语义网的目标是更为适度的：语义网并不是要构建一个通用的、综合性的、基于Internet的智能系统，而是要实现Web上数据集（datasets）间的互操作（无论数据是结构化、非结构化还是半结构化的）——这一目标更具实践性，更为适度。去年七月，Tim Berners-Lee专门就人工智能



Frank van Harmelen 是荷兰阿姆斯特丹自由大学人工智能系知识表示与推理研究组教授。W3C 推荐标准《OWL Web 本体语言》的制定者之一，开发了 Sesame（一种著名的 RDF 存储与检索引擎）的语义网领域元老级公司 Aduna 公司的科学顾问。已发表超过 100 篇论文，还有许多被广为引用（Hirsch 指数达到 35）。著有 5 本书，其中有一本关于语义网技术。第 15 届欧洲人工智能会议（ECAI2002）程序委员会主席，第 3 届国际语义网大会（ISWC2004）会议主席，以及第 14 届国际万维网大会（WWW2005）语义网专题主席。

与语义网之间的混淆做过一个报告。该报告的摘要如下：语义网不是人工智能，人工智能也不是语义网；人工智能是一个领域，而语义网是一个项目；语义网采用了人工智能里的某些工具，所以人工智能对语义网有一定功劳；语义网

将是人工智能的一个极好的应用场景。Tim Berners-Lee 还在报告中澄清了关于语义网的其他一些错误说法，例如“语义网（只或者主要）关注于手工标注的文本文档”以及“语义网需要一个可被大家接受的、统一的本体”等。

记者：Web 2.0 是一个新事物——无论是学术界还是工业界，人人都喜爱它。而另一方面，语义网却由于众多诺言未能兑现而失去关注。关于这两个 Web 的共存，您有何看法？您认为 Web 2.0 将对语义网的发展起到什么样的作用？

Frank：注意问题中的“语义网由于众多诺言未能兑现而失去关注”，这是一个错误的前提。

我们来看一些准确的信息：

SemTech 大会（Semantic Technology Conference）是一个面向工业界的会议，目前为止已经召开过 3 届，前几届都是在加利福尼亚圣何塞（San Jose）召开的。第一年有 300 人参加，去年有 500 人参加，而今年的参会人数已经超过了 700 人。相应地，在欧洲，首届欧洲语义技术大会（European Semantic Technologies Conference）也于去年 5 月在维也纳召开了。参会人数超过了 200 人，其中 75% 都是来自公司的。所以，要么你说错了，要么那几百名公司人士和几十家公司都“脑袋坏掉了”。你自己判断吧。

与此相反的是，语义技术正处于产业突破（Industrial Breakthrough）的过程之中。下面的一段话引自最近（2007 年 5 月）的一篇 Gartner 报告（Gartner

语义网不是人工智能，人工智能也不是语义网；人工智能是一个领域，而语义网是一个项目；语义网采用了人工智能里的某些工具，所以人工智能对语义网有一定功劳；语义网将是人工智能的一个极好的应用场景。

是不会热衷于短命技术的)：“关键结论，在接下来的10年里，基于Web的技术，将会提高往文档里嵌入语义结构、以及创建结构化词汇表和本体（用于定义术语、概念及关系）的能力。这将极大地推动信息的可见性及信息的开发利用（尤其是在无人工干预下理解文档与推断含义的能力）的发展。”

幸运的是，Gartner相当明智，没有（像你一样）断言语义网已经夭折，而是对语义网需要多久实现做出了估计：“实现语义网的伟大宏图，需要多个渐进式的步骤，而且小规模行动常常是最好的起点。”

回到刚才问题的主题上：学术界关于Web 2.0与语义网（或者说Web 3.0）是相互补充、而不是竞争的关系，取得了广泛的认同。这是在去年5月于爱丁堡召开的第15届国际万维网大会（WWW2006）^②上的一个科学座谈会上得出的结论。大家一致认为：Web 2.0门槛较低（易于上手使用），不过能力上限也比较低（folksonomy仅限于此）；而Web 3.0虽然门槛较高（起初的投入需要较多），但能力上限要高得多。

在前面提到的那份Gartner报告里，还有一些重要论断。它建议将语义网和Web 2.0技术结合起来，并预言目前的Web，将由具备轻量语义但易于使用的Web 2.0技术，逐渐发展为较高投入/较高产出的Web 3.0技术。

记者：您认为用自动化方法——也就是说，知识抽取与推理（Knowledge Extraction and Inference）中的自然语言处理（Natural Language Processing）、文本挖掘（Text Mining）及统计等方法——

来学习本体及实体间的关系怎么样？您认为这些技术对手工编辑本体的方法，是一种补充，还是遏止？您认为这些技术能够起到促进作用吗？还是它们注定会失败？

Frank：我对于本体的获取以及对这些本体里的数据对象的分类的态度是：有用就好。仅依靠手工来构建本体，显然是成本过高的，而且也限制了可被编纂与分类的知识量。所以，我希望你提到这些技术能为语义技术做出最大的贡献。我不觉得这些技术将“注定失败”，相反，我认为它们的有益贡献将日益增大，关于这一点我相当乐观。

记者：所有重大的技术发明与里程碑都是以其 Killer App 的出现为标志的。那么语义网的 Killer App 会是什么？究竟有没有？

Frank：我觉得关于“Killer App”这一反复问及的问题总是有些幼稚。举个例子：我们同不同意XML的广泛采纳是一项重要的技术创新？但谁是XML的“Killer App”？这样一个“Killer App”存在吗？不。存在的只是许许多多的场合，XML在这些场合下能够“在背后”起到促进作用。语义网技术主要是基础设施技术（Infrastructure Technology），而基础设施技术是存在于背后、并非用户直接可见的。你所能感受到的，只是诸如：网站变得更人性化了（因为背后存在语义网技术，你的个人兴趣概要（Profile）可以与网站的数据源进行互操作）、搜索引擎对结果的聚类（Clustering of Results）做得更好了（因为搜索引擎在背后采用具有含义的本体来对搜索结果加以分类）以及桌面搜索工具能够

把文档作者的姓名与你的地址簿里的Email地址关联起来了（因为这些数据格式在背后通过暴露其语义来实现互操作）等，但这些应用不会在其界面上注明“语义网技术”。语义网技术就像是发动机汽缸壁上的Nikasil涂层^③：虽然很少有司机知道它，但司机们能够察觉到燃料消耗的减少、最高时速的提升以及引擎寿命的延长等等。语义网技术就是目前正在开发着的、下一代人类友好的计算机应用的Nikasil涂层。■

译注：

①OWL（Web Ontology Language，Web本体语言）2004年成为W3C推荐标准，是目前最常用的一种本体描述语言。《W3C OWL Web Ontology Language Guide》（<http://www.w3.org/TR/owl-guide/>）是一个很好的入门材料，其中译本位于<http://zh.transwiki.org/cn/owlguide.htm>。

②国际万维网大会（International World Wide Web Conference）是Web领域的顶级会议，在学术界和工业界均享有极高的声誉，目前已召开16届。下一届国际万维网大会（WWW2008）将于2008年4月在中国北京召开。

③Nikasil涂层是Yamaha公司发明的一种专利技术。Nikasil是一种镍矽碳化物，将它电镀在汽缸上可以解决铝硅发动机耐磨性问题。该技术一般在应用在高档跑车上。

译者简介

徐涵，中文W3C技术推广网站W3China（w3china.org）创始人，开放翻译计划（transwiki.org）发起人。2005年毕业于东南大学，获计算机硕士学位。2006年翻译出版国内第一本SOA专著《Understanding SOA with Web Services 中文版》。研究兴趣包括：Web Architecture、Semantic Web、Web Services、SOA、Social Network等。

■ 责任编辑：刘龙静（llj@csdn.net）