

# 采用CIDOC概念参考模型为基础的语义维基 (semantic wiki)进行终端用户故事讲述

Vincent Ribaud (Patrick Le Boeuf<sup>1</sup>编审)

## 1. 导论

本论文讲述一群用户在采用CIDOC参考概念模型作为知识表达语言上的现有经验。我们把大一STEM（即科学、科技、工程和数学）学生分成几组，一组有2至4个成员，然后要他们选择一个主题。每组必须在一个网络承载的语言维基中建模、建构、编写和呈献故事。CIDOC参考概念模型的主要部分被设为本体核心，学生们可在此附加和故事相关的领域分类和属性。我们做了2项假设：其一，CIDOC参考概念模型能提供稳定和可扩展的基础；其二，参考概念模型能自然而顺畅地引导终端用户阅读和编写故事。

## 2. 工作假设

综观我们所作出的选择（以及可能获得的效益），我们可以这么说：第一，我们从各种知识管理系统当中选择了语义维基，主要是因为它使用方便，容易上手。第二，我们会选择MediaWiki及其语义扩展(semantic extension) SMW (Semantic MediaWiki [Kr. tzscho7])，主要是维基百科(Wikipedia)的普遍流行和使用MediaWiki（以及SMW）的系统很多，因此大家都对它们比较熟悉。第三，从各种合适的本体当中，我们会选择CIDOC概念参考模型[Crofts and al. 2010]，主要是因为它范围广泛、能自由扩散，而且富有存档和呈献等相关资源。我们的选择并非基于认识论或试验性原因，而是纯粹以“简单最好”的原则出发—这显然会缩小论文的科学范畴。

## 3. 语义维基和CIDOC概念参考模型

维基基本上是一系列交叉链接的网页，而一个语义维基允许我们为各个链接赋予不同的意义，并对一套与每个网页有关的属性赋值（通称元数据）。在Media Wiki网页上所使用的陈述语法(statements syntax)含有潜在语义。它根据网页所属于的分类提供主要指示。因此熟悉定义域的用户可以期待找到某些与属性相关的价值，并用于这些分类上。在一个语义维基里，这些价值（以及价值所实例化的属性）由终端用户进行形式界定、在网页出版后被分析、以RDF三元组(RDF triples)处理和存储，以及在语义检索中被取回。

<sup>1</sup>需征求Patrick Le Boeuf的同意。

共同模式（本体）选择允许用户社群分享意义，而我们选用一个根据“缩减的参考概念模型兼容形式”（reduced CRM-compatible form）的扩大版本为本体核心，有关版本的详情请参照“CIDOC概念参考模型的定义” [Crofts and al. 2010]。我们另外增添了12个分类（以及众多的相关领域属性），新内容主要是E21 人物、E40 法律机构、E31 文件、E38 图像，以便实现“本体闭包”（“ontological closure”）。

## 4. 课程目标和内容

此课程在过去2年里已经开办了4次：2次在信息科技硕士班开办，另2次在一年级STEM（即科学、科技、工程和数学）学生的学士班开办。我们在课程第2版中大副增加上课时间，并大量削减内容。现有课程版本的上课时间为15小时（这包括中间休假但必须做作业的时间），另外还有12小时的实验时间。导师会按照学生的要求，无限量提供辅导时间，并通过交换电子邮件和介入学生的维基指导学生。

硕士班对第1版本的课程进行评估后，我们便尽量把Web 2.0 技术的繁复细节隐藏起来。一年级学生对第1版本课程的评估，则促使我们决定采用CIDOC 概念参考模型的精简版作为本体核心，并以实体-关系(Entity-Relationship)表示法而非语义网络表示法支持知识表达的建模。

为一年级女生所开办的第2版课程中，我们从定量（有赖维基统计数字）和定性（根据设定目标对每个维基的评估）两方面进行评估和衡量。在为课程内容和目标分类进行简单介绍之后，本论文的其余部分将尝试对定量和定性结果进行诠释，以便识别出课程失败之处，研究改进方法。

## 5. 目标的分类和种种困难

课程目标可分为4级。此分类法更接近ISO 15504 水平的“能力水平”（capability level）概念，这是布鲁姆分类(Bloom's taxonomy) [Bloom56]等分类法所不及的。15504标准以由6个流程功能阶层以及相关流程属性组成的测量框架为根据，具备了能标出“能力”的特点[15504, Part 2, p. v]。在15504标准里，等级分类测量法被用于测量流程属性的实现程度[15504, Part 2, pp. 1011]。我们采用的是相同的目标分类量表，测量有关目标在语义维基里的应用情况：N Not achieved（没有实现），P Partially achieved（部分实现），L Largely achieved（大体实现），F Fully achieved（完全实现）。上述定义的有序点(ordinal points)将根据表示实现程度的百分率量表进行解读。

下列为每个阶级的环球和详细目标：

### 1. 第1级: 复制 (*reproduction*)

学生能够丰富维基数据(内容)但是很少能够充实其结构。此阶段的目标是:

1. 理解MediaWiki和SMW的基本原则和语法。
2. 理解分类(类别)和网页(实例)之间的区别。
3. 理解属性和关系之间的区别。

### 2. 第2级: 定制 (*customization*)

学生能够丰富维基内容和充实其结构。此阶段的目标是:

1. 理解属性量化对用户界面和维基结构的影响
2. 理解什么是定义域以及属性值域和使用后果。
3. 界定和实行语义检索。

### 3. 第3级: 变换 (*transformation*)

学生能够使用概念参考模型为故事域建模,并在语义维基内推行。此阶段的目标是:

1. 理解n元关系(n-ary relationship)和二元关系的实行。
2. 理解什么是以事件为中心的建模形式。
3. 界定和实行相反属性(inverse attributes)(尤其是通过语义检索)

### 4. 第4级: 检讨 (*reflection*)

学生在语义网上达到真正良好的能力水平。此阶段的目标是:

1. 理解“本体方形”(ontological square<sup>2</sup>)。
2. 理解“属性的属性”以及其实行上的问题。
3. 认识多语言等互操作问题。

## 6. 维基评价

### 5. 中介评估 (*Intermediary appraisal*)

课程结束前2周我们进行评估。图表1列出了定量事实(quantitative facts)。

下列为图表的各列标题: 2 - 小组的实际学生人数;

3 - 小组内的女生人数百分比; 4 - 加入概念参考模型核心的类别数目; 5 -

加入概念参考模型核心的被使用属性数目; 6 -

赋值属性数目(概念参考模型和特定域); 7 - 上传文件数量; 8 -

维基访问次数。列号1列出了维基的网站链接(URLs),但都以法文书写,因此对不谙法语者可能有些不便。

图表1: 定量统计

---

<sup>2</sup>本体方形是一个具备4个分类的模型,通过交叉类型(或共相)和个例(或殊相)之间、以及字符(或特征)和其载体(或基板)的两个形式区别[...]获得实现[Schneider 2008]。

ABSTRACT

Wiki URL	学生人数	女性 %	类别数目	模板数目	被使用属性数目	赋值属性数目	上传文件数量	维基访问次数
Alien2.referata.com	3	33%	3	3	13	200	9	679
Charlietlachocolaterie.referata.com	3	67%	6	13	29	547	19	1618
Clamp.referata.com	2	100%	5	5	30	530	33	1385
<i>Finalfantasy13.referata.com</i>	<i>1</i>	<i>0%</i>	<i>2</i>	<i>5</i>	<i>11</i>	<i>211</i>	<i>22</i>	<i>470</i>
<i>Lesfrerescoen.referata.com</i>	<i>1</i>	<i>0%</i>	<i>0</i>	<i>0</i>	<i>1</i>	<i>3</i>	<i>0</i>	<i>55</i>
<i>Lesroismaudits.referata.com</i>	<i>1</i>	<i>0%</i>	<i>0</i>	<i>0</i>	<i>1</i>	<i>26</i>	<i>0</i>	<i>127</i>
Lol.referata.com	4	75%	3	3	9	185	19	578
<i>Lost.referata.com</i>	<i>1</i>	<i>0%</i>	<i>1</i>	<i>1</i>	<i>4</i>	<i>32</i>	<i>0</i>	<i>142</i>
Masseffect.referata.com	2	0%	6	2	14	182	1	551
Michaeljackson.referata.com	2	0%	6	5	30	378	10	798
Ncis.referata.com	3	33%	4	3	30	51	0	575
<i>Oasis.referata.com</i>	<i>1</i>	<i>0%</i>	<i>5</i>	<i>4</i>	<i>14</i>	<i>125</i>	<i>4</i>	<i>156</i>
Skins.referata.com	2	50%	2	0	24	61	14	785
Stargatesatlantis.referata.com	2	0%	3	3	18	222	8	550
Stargatesg-1.referata.com	2	0%	3	3	8	41	1	252
Starwars.referata.com	2	0%	3	5	15	143	0	436
Tintin.referata.com	4	100%	4	4	32	279	18	1137
<i>Warcraft3.referata.com</i>	<i>1</i>	<i>0%</i>	<i>1</i>	-	1	-	<i>102</i>	<i>2568</i>

有些小组只有1个学生（课程一开始或因学生中途退学），我们以斜体字呈献和他们相关的数据。我们使用第5分段的阶级和目标定性评估，在对18个语义维基（由37个积极参与的学生负责）进行评价后发现：

- 1个维基处于无人照管的状态；
- 1个维基“大体”或“完全”达到第1级；
- 6个维基“大体”或“完全”达到第1级，但是没有达到第2级；
- 5个维基“大体”或“完全”达到第2级，但是没有达到第2级；
- 5个维基“大体”或“完全”达到第3级，但是没有达到第4级；

我们也发现有2到5名学生凭个人力量达到第4级。

维基评估活动让我们发现2项主要问题：

第1个问题：对非具体工作（与FRBRoo的F2表达 [Expression] 的概念相应）和（具体）实现（与FRBRoo的F5项目 [Item] 的概念相应）之间感混淆

第2个问题：这项困难（这在某些维基里与定义域主题有关，例如电子游戏）让以事件为中心的建模法（也就是说以一种实体为本的持续性视角建模）很难实现。

为了解决第1个问题，我们赶紧增设了一堂课，介绍FRBRoo的简易版，并指示同学们如何更新维基，以便分辨F2表达元素和F5项目元素之间的不同。我们目前正在积极解决第2个问题，但是至今无法找到一个全面的解决方案（即适用于所有故事：艺术家群体或个别艺术家、电影、电视系列、电子游戏和卡通片）。因此在现阶段仍然需要个别指导学生，给予他们特定指示，让他们能够把某些维基带入轨道(rails)。

## 6. 最后的评价（定量和定性）

最后的评估将需要2周完成。此项为中介评估，将包括定量（与维基统计数字有关）和定性（与所达到的阶级有关）。我们也将尽力尝试（局部或全部）解决上述问题。

## 7. 其他结果

我们可以根据经验指出，在表现令人满意的小组当中，若论其性别构成，那些有女生参与或清一色是女生的小组，所获得的评价一般比清一色男生组成的小组高出许多。此项目所可能取得的第2项结果—必须和其他同类结果相比较，例如[Hardy 08]—便是一个依赖Web

2.0科技的课程，我们也会采用小组作业的教学形式，因为这比传统的编程课程更能够吸引女生。

值得注意的是，学生们采用电子邮件和维基，与导师们交流时可谓如鱼得水，这可能是由于他们都属于Y世代或网络世代的关系<sup>3</sup>。

## 8. 结论

我们的目的是让STEM学生了解，计算机科学并不只是和土包子才稀罕的计算机编程有关。我们的意图之一是吸引女生进入计算机研究科。互联网、电子商务、多媒体、社会知识创建等日常生活应用程序都被设为起点，以便引导学生进入计算机课题。但看过学生在大二时候对课程的选择后，却显示这种“宣传策略”并没有产生预期中的效果。然而在本文所介绍的课程个别部分，的确引起了学生的兴趣。课程总算取得一定的成功，一般而言能对本体和语义网的推广作出贡献，具体而言则对CIDOC概念参考模型的普遍推行能起一定作用。

## 9. 参考文献

---

<sup>3</sup> [http://www.businessweek.com/managing/content/mar2008/ca20080313\\_241443.htm](http://www.businessweek.com/managing/content/mar2008/ca20080313_241443.htm)

## ABSTRACT

Maedche, A. ; Motik, B. ; Stojanovic, L. ; Studer, R. ; Volz, R. ; 2003, Ontologies for enterprise knowledge management, Intelligent Systems, IEEE , vol.18, no.2, pp. 26-33, Mar-Apr 2003  
doi: 10.1109/MIS.2003.1193654

Kr.tzsch, M. ; Vrandecic, D. ; V.lkel, M. ; Haller, H. and Studer, R. ;  
“Semantic Wikipedia” ,  
Journal of Web Semantics 5/2007, Elsevier, Amsterdam, 2007, pp. 251 - 261.

Crofts, N. ; Doerr, M. ; Gill, T. ; Stead, S. ; Stiff, M. (editors),  
“Definition of the CIDOC  
Conceptual Reference Model” , January 2010,  
[http://cidoc.ics.forth.gr/docs/cidoc\\_crm\\_version\\_5.0.2.pdf](http://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.2.pdf) (last  
accessed, April, 39th, 2010)

ISO/IEC 15504:2004, “Information technology --Process assessment” .  
Geneva: ISO, 2004.

Benjamin S. Bloom & David R. Krathwohl, The classification of educational goals, by a committee of college and university examiners, Longmans, New York, 1956

Schneider, L. 2008. The Ontological Square and its Logic. In Proceeding of the 2008 Conference on Formal ontology in information Systems: Proceedings of the Fifth international Conference (FOIS 2008) C. Eschenbach and M. Grüninger, Eds. Frontiers in Artificial Intelligence and Applications, vol. 183. IOS Press, Amsterdam, The Netherlands, 36-48.

Hardy, N. 2008. Women in computer science: harnessing the power of web 2.0 to draw women to computer science fields. In Proceedings of the 9th ACM SIGITE Conference on information Technology Education (Cincinnati, OH, USA, October 16 - 18, 2008). SIGITE '08. ACM, New York, NY, 59-60. DOI= 10.1145/1414558.1414576