

## DH导论

课堂案例分享

由中国人民大学信息资源管理学院冯惠玲教授、刘越男教授、严承希博士和哲学院王小伟副教授主讲的《数字人文导论》是中国人民大学本科生通识核心课程之一，面向全校本科生开设。课程内容包括数字人文的基本概念、历史源流、发展脉络、基本方法和技术、前沿议题等。本系列推文是该课程学生的学习成果展示。

本期作者

许麒 公共管理学院2020级

明菁 财政金融学院2020级

秦澄宇 明德书院2021级

“中国陶瓷”项目案例评介

## 壹

案例引入



“中国陶瓷”项目（Chinese Ceramics，项目网址：<http://www.dh.ketrc.com>），曾获2020年数字人文奖“最佳数字人文数据集”提名奖，是一个致力于定义中国陶瓷器皿术语概念的跨学科合作项目。该项目旨在为博物馆提供开放数据发布模型，链接中国陶瓷数据，构建中国陶瓷多语言电子词典，与世界考古学家交流。

## 贰

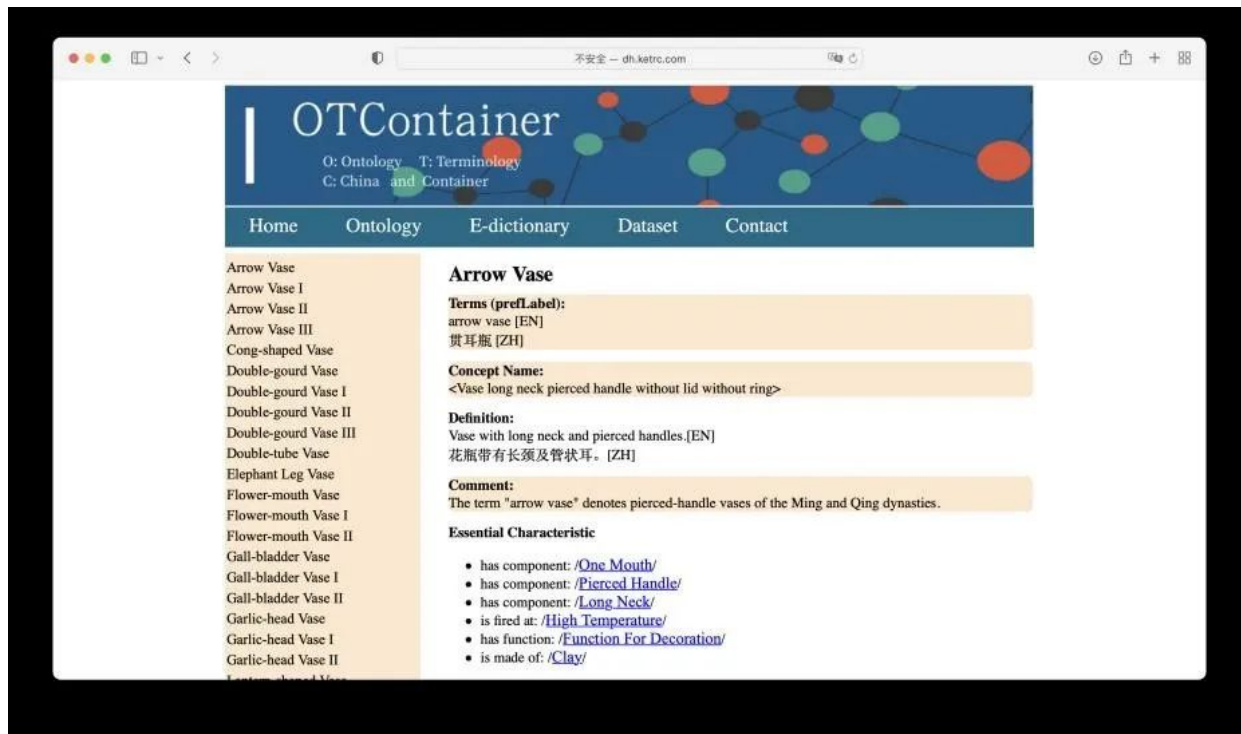
项目背景

### 01 研究背景

文化遗产是指一个群体或社会从过去若干代人那里继承下来的实物物品和无形属性的遗产，而术语则是文化遗产传播的工具。当前国内文化遗产保护工作日臻成熟，但面向国际的交流互鉴明显不足，国际认可度有待提升，数字人文研究的介入，有助于扫清国际文物保护交流障碍。

同时，中国各地的博物馆里收藏着大量的中国陶瓷花瓶，可以为历史研究提供丰富的数据来源。

## 02 研究目标与意义



该项目旨在通过实现语义网和ISO术语标准(ISO 704和ISO 1087-1)的实践来应对跨语种下的文化遗产保护工作者的交流沟通与工作互认问题，并将结构化数据结果发布到语义Web上以供任何感兴趣者使用。该研究开拓了中国陶瓷全面本体构建的先河，为中国文化遗产其他领域的本体构建提供了借鉴，寄托着项目团队对于帮助越来越多的中国遗产机构发布开放的文化遗产数据并进行链接的美好希冀。

## 03 团队介绍

### 【团队简介】

该项目团队以中国聊城大学和法国萨瓦勃朗峰大学于2017年共同成立的KETRC（知识工程与术语研究中心）为基础，项目负责人为克里斯托夫-罗什教授。

### 【项目负责人简介】



## Home

- Full Professor in Artificial Intelligence, University Savoie Mont-Blanc (France)
- Special Appointment Professor, University of Liaocheng (China)
- Head of the [Condillac Research Group](#) on « Terminology & Ontology » – LISTIC Lab. University Savoie Mont-Blanc (France)
- Dean of the [KETRC](#) Research Center on « Knowledge Engineering and Terminology », University of Liaocheng (China)
- Researcher & Lecturer at University NOVA of Lisbon – CLUNL Linguistic Research Center (Portugal)
- PhD in Artificial Intelligence (INPG Grenoble – 1984)
- Domains of interest: Artificial Intelligence, Linguistics, Digital Humanities, Knowledge Representation, Terminology, Ontology, Ontoterminology
- Chairman of the [TOTH Conferences](#)
- Chairman of the AFNOR Commission on Terminology ([X03A](#))
- Project Leader of the ISO Standard on Terminology (ISO 1087:2019)
- 2017 Talent Foreign Expert (Shandong Province)
- 2018 “Qilu Friendship” Award (Shandong Province)

Christophe Roche, 法国萨瓦大学教授，博士生导师，里斯本新大学副研究员，葡萄牙语言研究中心研究小组组长，法国人工智能知名学者，主持欧洲FP7项目多项，语料本题库国际专家，负责制定术语方法准则704和860标准。

### 【团队成员】

Tong Wei (China)

Yangli Jia (China)

Papadopoulou Maria (Greece)

Julien Roche (France)

### 【合作机构与赞助方】

该项目有两个合作团队，分别是萨瓦勃朗峰大学LISTIC实验室的康迪拉克小组以及聊城大学计算机学院。同时，该项目在2017年11月至2020年11月之间曾获得来自国家留学基金委的资助。

## 叁

### 项目实施过程

本体构建遵循几个阶段，其中一些阶段是普遍的，另一些阶段因为考虑到该领域的特殊性而必须额外加入。支撑本体论的概念理论也会对构建方法产生强烈影响。遵循ISO术语原则，其中“术语是概念的口头表示”，“概念是（基本）特征的独特组合”，他们采用了一种“术语和特征”指导方法，该方法源自数字人文学科的工作。

识别基本特征成为主要目标。这一阶段的目的是识别物体之间的差异（有颈花瓶和无颈花瓶），并对汉字进行形态学分析，这些汉字与所表示的物体有关。

术语和特征指导方法包括七个步骤。每一步都针对不同的任务：

第一步：确定领域范围和目标；

第二步：确定术语和对象；

第三步：确定基本特征；

第四步：定义概念；

第五步：使用可用工具之一构建本体；

第六步：整合其他资源；

第七步：评估。

## 肆

### 技术方法或工具

这里介绍的最新技术的特点是，为文化遗产数据的表达制作了可互操作的词汇表。它包括W3C语言、语义数据模型、同义词库和本体资源，以及本体构建方法。

**W3C标准:**在web上开放文化遗产依赖于W3C标准，如OWL 12（web本体语言），一种语义web语言，旨在表示有关事物、事物组和事物之间关系的丰富而复杂的知识；SKOS13，一种用于在web上共享和链接知识组织系统的数据模型。SKOS可以用来获取博物馆和其他记忆机构现有主题词表的大部分语义。他们也引用DC，这是一个基于15个基本属性的元数据模式，用于描述在线和物理资源。



**CIDOC-CRM:**考虑文化遗产领域的语义数据模型。尤其是CIDOC-CRM，这是一种元本体，用于表示博物馆和文化遗产专家使用的概念（CIDOC，2003）。它提供了一个语义框架，用于在不同的文化遗产资源之间建立映射，减少其异质性（Doerr，2003）。

**AAT:**至于陶瓷项目可以受益的本体论资源，引用AAT（艺术与建筑辞典），这是一种结构化资源，可以通过丰富的元数据和链接来改善对艺术、建筑和其他物质文化信息的访问，与其他词汇一起为数字艺术史和相关学科的研究和发现提供强大的渠道。AAT包含超过250000个关于建筑历史、风格和技术的术语。本体已与AAT链接，以便为本体中的术语提供更多信息。

**Kerameikos:**这是一个致力于定义陶器的知识概念的合作项目（<http://kerameikos.org/>），遵循链接开放数据的原则，并制定一个本体，用于在不同的数据系统中表示和共享陶瓷数据。

**NeOn:**这是一种方法（Suárez Figueroa等人，2015年），提供了开发本体论的九种场景。

## 伍

### 项目成果

陶瓷本体包含165个类、11个对象属性、8个数据属性、132个个体和3124条公理。它与CIDOC CRM和AAT映射。

所有物的对象属性包括belong To（领域：皇帝，范围：王朝）、has Function（领域：容器，范围：功能）、has Component（领域：容器，范围：组件）等。

## 陆

### 项目评析

#### 01项目特点

- a. 跨学科属性：该项目涉及诸多学科，包括考古学（文化遗产）、语言学（术语）、人工智能（知识表示）、语义网（链接数据）等。
- b. 独特目的：不是简单地对中国陶瓷的信息进行集成整合，而是对陶瓷进行本体构建，并使之与其他文化遗产资源之间建立映射。
- c. 题材特殊：对于我国宝贵的非物质文化遗产——陶瓷进行了本体构建和数字化，对于陶瓷文化的数字化保存和传承有着重要意义。
- d. 工具特殊：使用了W3C语言、文化遗产领域的语义数据模型和艺术与建筑辞典(AAT)，旨在在不同数据系统中最大程度地表示和共享陶瓷数据。

e. 国际合作：本项目由中国KETRC（知识工程与术语研究中心）完成，与萨瓦勃朗峰大学 LISTIC 实验室的康迪拉克小组、聊城大学计算机学院合作。

## 02项目评价

| 维度↵  | 指标↵   | 说明↵                 |
|------|-------|---------------------|
| 学科性↵ | 学科资料↵ | 权威、系统↵              |
|      | 数据↵   | 专业、有效↵              |
|      | 项目成果↵ | 实用↵                 |
| 专业性↵ | 项目资源↵ | 资源广泛，数据关联性强↵        |
|      | 数据保存↵ | 完整、易于获取↵            |
| 技术性↵ | 数字方法↵ | 遵循 ISO 术语原则，数据语言标准↵ |

## 03项目启示

### a. 文化遗产的数字化留存

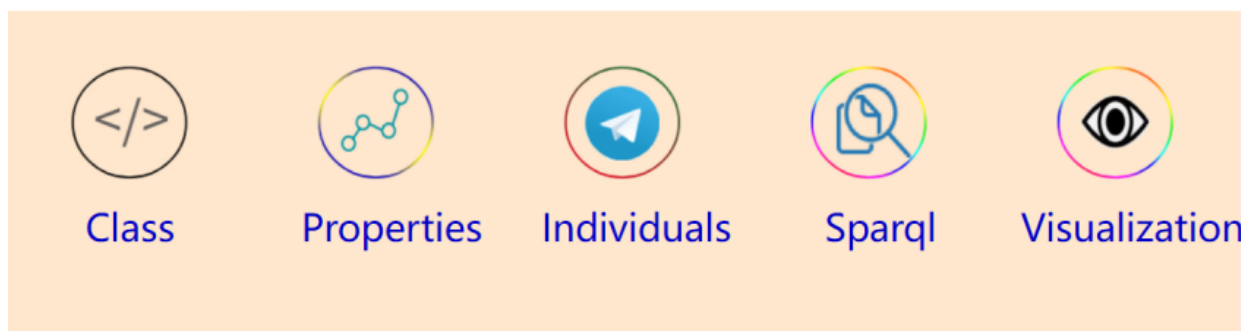
该项目题材本身对于公共文化遗产的数字化留存具有意义。非物质文化遗产的保护和继承本身难度大，通过该数字人文项目，能够为后世留下系统的关于中国陶瓷的信息。

### b. 多学科交叉的优越性

该项目不仅仅从考古学来认识文物，更借助语言学、人工智能、语义网来重新定义陶瓷器皿术语，并建立电子辞典，以便于与世界各地考古学家进行交流。

## 04批判性分析

### a. 界面设计及功能使用



在网站的Ontology界面，这些具体的查询检索功能的图标不能点击跳转，不符合人们的网络使用习惯。且网站中并没有各项功能的详细介绍和使用方法，给用户的使用带来了困难

### b. 数据可持续性

该项目的数据可持续性较低，缺少数据的持续更新和补充。

根据该项目在2020 KEOD会议的报告，项目团队的下一步工作是丰富语言维度，从“类”进行扩展，并考虑加入其他类型的陶瓷容器数据来进一步完善本体论。

排版：高宇博



# 数字人文研究

左 数 字 ， 右 人 文

