

《智能信息处理》课程考试

基于本体的连衣裙款式检索研究

张紫英

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2021 年 12 月 19 日

基于本体的连衣裙款式检索研究

张紫英

(大连海事大学 信息科学技术学院, 大连 116026)

摘要 随服装款式丰富多样, 服装属性语义复杂。针对关键词难以表述消费者检索需求的问题, 文章以款式复杂的连衣裙为例, 提出了一种基于本体的连衣裙款式知识库构建方法, 对其他类别服装语义检索具有普遍适用指导意义。通过汇总和归纳连衣裙款式概念及其关系, 建立款式关键词的关联性, 构建连衣裙款式本体模型。并结合本体语义扩展方法, 提出了基于语义信息的连衣裙款式检索模型, 实现基于语义的连衣裙款式智能搜索, 提高款式检索的查全率和查准率。

关键词 本体知识库; 连衣裙款式; 相似度; 语义检索; 语义扩展

Research on Dress Style Retrieval Based on Ontology

Zhang Ziying

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

Abstract With the various styles of clothing, the semantics of clothing attributes are complex. Aiming at the problem that keywords are difficult to express consumers' retrieval needs, the article takes dresses with complex styles as an example, and proposes an ontology-based dress style knowledge base construction method, which has universally applicable and instructive significance for semantic retrieval of other categories of clothing. By summarizing and summarizing dress style concepts and their relationships, establishing the relevance of style keywords, and constructing dress style ontology models. Combined with the ontology semantic extension method, a dress style retrieval model based on semantic information is proposed, which realizes the intelligent search of dress styles based on semantics, and improves the recall and accuracy of style retrieval.

Key words Ontology knowledge base; Dress style; Similarity; Semantic retrieval; Semantic extension

1 引言

形随着服装领域信息化的发展, 消费者通过网络就能进行服装的查询, 服装款式检索成为服装行业研究的重点。服装款式具有服装品类多、款式丰富、服装属性语义复杂等特征, 但目前服装领域的检索引擎大多是基于关键词检索的方式, 而简单的关键词检索无法在语义层面通过几个关键词完全理解用户特定的需求, 信息检索结果通常广泛而不精确, 因此将信息检索系统进行

细分, 针对专业领域构建信息检索系统是近几年的研究趋势。

本体作为一种建模工具, 提供了一种恰当的概念表示方式, 通过构建语义模型, 充分地表达领域内概念的层次结构和概念间的关系, 精确地描述领域内的知识。本体在语义信息检索系统的语义标注、基于本体的索引、基于本体的查询扩展等多个方面都有很重要的作用。

连衣裙因春夏可单穿, 秋冬可内搭为四季常青类目, 且连衣裙款式复杂, 覆盖多种廓形、领型等款式类型, 具有广泛代表性。因此本文以连衣裙为例, 提出了基于语义层面上的连衣裙款式

语义检索模型，对其他类别服装的语义检索具有普遍适用指导意义。通过分析连衣裙款式知识及概念层次，利用连衣裙款式领域概念之间的语义关联，构建连衣裙款式本体模型。并结合本体语义扩展算法，将本体数据库在连衣裙款式信息检索中进行语义推理应用，实现连衣裙款式语义检索。

2 本体概念及相关理论

本体(Ontology)的概念最初起源于哲学领域，可以追溯到公元前古希腊哲学家亚里士多德。它在哲学中的定义为“对世界上客观存在物的系统地描述，即存在论”，是客观存在的一个系统的解释或说明，关心的是客观现实的抽象本质。

在人工智能界，最早给出 Ontology 定义的是 Neches 等人，他们将 Ontology 定义为“给出构成相关领域词汇的基本术语和关系，以及利用这些术语和关系构成的规定这些词汇外延的规则的定义”。Neches 认为：本体定义了组成主题领域的词汇表的基本术语及其关系，以及结合这些术语和关系来定义词汇表外延的规则。后来在信息系统、知识系统等领域，越来越多的人研究 Ontology，并给出了许多不同的定义。其中最著名并被引用得最为广泛的定义是由 Gruber 提出的，本体是概念化的明确的规范说明。

本体是共享的概念模型的形式化的规范模型，Fensel 认为主要包括 4 个方面：

- 1) 概念化：客观世界现象的抽象模型；
- 2) 明确：概念及概念之间的联系都被精确定义；
- 3) 形式化：精确的数字描述；
- 4) 共享：本体中反映的知识是其使用者共同认同的。

依据本体表达的概念模型所在的层次，可对本体做如下分类：知识本体，通用本体，领域本体，应用本体。各种分类也应用不同的研究领域。

3 连衣裙款式领域本体设定原则及流程

连衣裙款式本体是定义连衣裙款式领域、可以被计算机理解且被该领域所共同接受的一些概念，

通俗地说，是建立检索连衣裙时所用的正式的词汇表，包含了连衣裙款式领域内的概念、实例、概念之间的关系及概念的属性等。领域本体结构如图 1 所示。

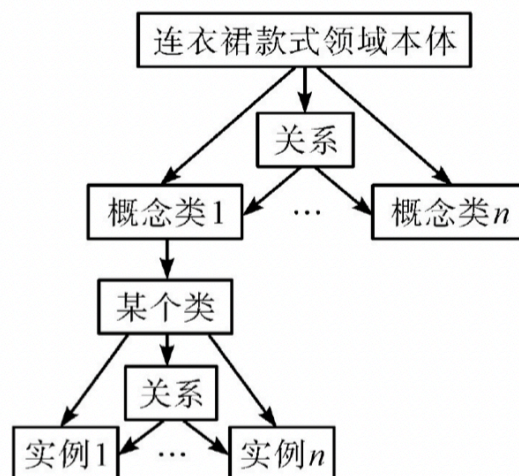


图 1 领域本体结构

连衣裙款式本体在“七步法”建模方法基础上，结合服装产品的特点及款式检索的需求，形成的具体构建过程如下：1) 确定连衣裙款式本体的领域和范围，即本体构建的应用目标；2) 收集领域相关信息；3) 列举领域内最核心的概念，一般取自教科书或者专家知识；4) 核心概念词汇层次化；5) 核心概念同义词扩展；6) 确定概念之间的关系和属性；7) 利用本体工具构建连衣裙款式本体。

4 连衣裙款式本体的构建

本文构建的连衣裙款式本体主要面向检索应用领域，一般在款式检索时需要款式进行标注，而标注的过程就是获得描述款式的语义概念的过程，因此构建连衣裙款式领域本体最主要的任务是收集款式领域的概念词汇。服装款式领域本体概念的构成及分类，形成连衣裙款式领域本体的核心层次概念；另一方面通过对消费者问卷访谈进行概念同义词扩展。

服装款式领域的一个检索障碍是词汇的多样性，主要是连衣裙款式的廓形、领型、袖型，不同地区不同人的叫法不同，导致用户的检索词汇与本体中的概念语义描述不尽相同。为了解决用户检索同义类型词时信息不足的问题，需要对本体中的词汇进行同义词扩展。例如，“H 形”的同义关系概

念有“直身形、直筒形”，当用户在领域本体模型中检索“H 形”时，通过本体语义扩展，检索词将扩展为检索集“H 形、直筒形、直身形”，以此提高用户检索的准确性。部分关键词的同义扩展汇总如表 1 所示，连衣裙款式本体概念关系集如表 2 所示。

表 1 部分关键词的同义词汇汇总

专业术语	用户常用词汇	专业术语	用户常用词汇
H 形	直身形、直筒形	娃娃领	彼得潘领
A 形	瓶形、金字塔形	宝蓝色	克莱因蓝
X 形	修身形、沙漏形、喇叭形	喇叭袖	荷叶袖
O 形	茧形、圆球形、泡泡形	泡泡袖	公主袖
U 领	船领、方领	连身袖	中式袖、蝙蝠袖
立领	中式领	翻驳领	西装领
衬衫领	衬衣领	直袖	圆袖
海军领	水手领	中长袖	七分袖

表 2 连衣裙款式本体概念关系集

RS 集	关系表达	关系说明
组成关系	Part-of, has-part	概念之间部分与整体的关系
继承关系	Kind-of	概念间的继承关系
实例关系	Instance-of, has-instance	某个概念是另外一个概念的实例
属性关系	Attribute-of, has-attribute	某个概念是另外一个概念的属性
同义关系	owl: Equivalent to	概念之间的同义关系

5 基于本体的连衣裙款式检索

领域本体中概念之间的语义相似度是语义信息检索研究的重点。本体概念相似度计算是指在计算概念的相关度时参考了概念间的层次结构信息，如节点关系、语义邻居关系等。在款式检索中使用语义相似度来代表这些关系，连衣裙款式领域本体可以看成一本体层次概念语义树，如图 2 所示。其中结点 N_1 、 N_2 、 N_3 等表示本体中的概念词，结点之间的边表示本体中概念与概念之间的语义相似度，如 $S(N_1, N_2)$ 。

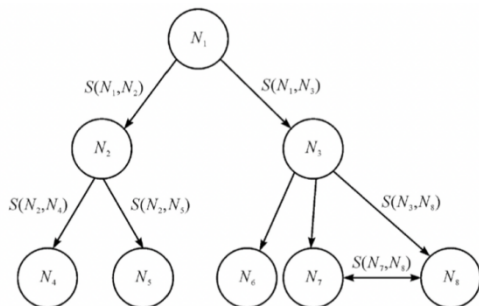


图 2 本体概念语义树

本文采用概念语义相似度的计算方法，用 $S(x, y)$ 表示两个概念之间的相似程度。如果连衣裙款式本体中概念 N_i 和 N_j 是同义关系，则两者的语义相似度为 1，即 $S(C_i, C_j)=1$ ；如果是两个非同义的概念 N_i 和 N_j ，则两者之间的语义相似度按下式计算。

$$S(C_i, C_j) = \frac{(Dist(N_i, N_j) + \theta) \times \theta(d(N_i) - d(N_j))}{CE(N_i, N_j) \times 2 \times Dep \times \max(|d(N_i) - d(N_j)|)} \quad (1)$$

式中， $d(N_i)$ 和 $d(N_j)$ 分别是概念 N_i 和 N_j 对应的节点在概念树中的层次， $d(N_i)-d(N_j)$ 表示概念 N_i 和概念 N_j 的层次差， $Dist(N_i, N_j)$ 表示从概念 N_i 结点到概念 N_j 结点所经过的路径上相似度的和。 θ 是一个可调节参数，用于调节概念层次差对语义相似度的影响，一般大于等于 0。 $CE(N_i, N_j)$

表示从概念 N_i 点到概念 N_j 结点所经过的路径长度，每条边记数为 1，当 N_i 和 N_j 不连通时，路径长度为 ∞ 。 Dep 为概念所在本体中的层次深度。

将建立的连衣裙款式知识本体结构连接到款式检索的系统中。基于本体的连衣裙款式检索流程如下：用户发出语义检索请求，首先关键词进入本体进行关键词扩展，包括在领域本体的基础上对用户的检索语言进行同义和相似扩展(如果是单关键词也直接进行语义相似度扩展)，生成新的语义检索扩展集合，然后在款式库中检索匹配的款式，最后将相似性款式检索结果推荐给用户。基于本体的连衣裙款式检索系统流程如图 3 所示。

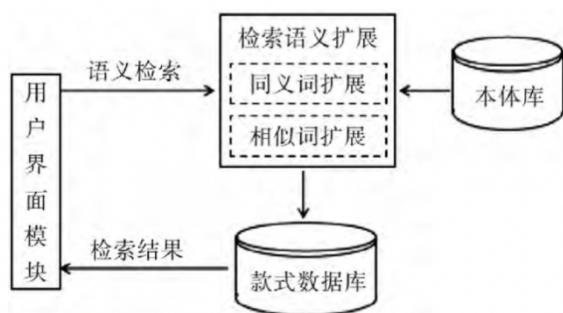


图 3 基于本体的连衣裙款式检索

目前常用两个指标来检验一个检索系统的检索效果，即查全率(Recall) 和查准率(Precision)。公式如下：

$$P/\% = n/N \times 100 \quad (2)$$

$$R/\% = n/M \times 100 \quad (3)$$

式中： n 指查询结果中与检索请求标注相同的款式数量， N 指查询结果中所有款式数量， M 指系统中所有相关的款式数量。

为了验证基于本体的款式检索模型的有效性，本文从淘宝、京东等常见购物网站选取 80 件已标注产品关键词的连衣裙款式，分别按传统简单关键词检索方式和基于本体的语义检索方式进行检索。通过结果比较分析，两种检索方式得到的查准率和查全率两项性能指标对比如表 3 所示。

表 3 两种检索方式查询结果对比

检索词	查全率/%		查准率/%	
	基于本体的检索	基于关键词的检索	基于本体的检索	基于关键词的检索
O 形	94.54	80.15	92.76	87.04
圆领	93.42	93.21	93.45	89.73
羊腿袖	94.57	94.46	91.08	89.60
方领	96.39	72.18	90.57	76.92

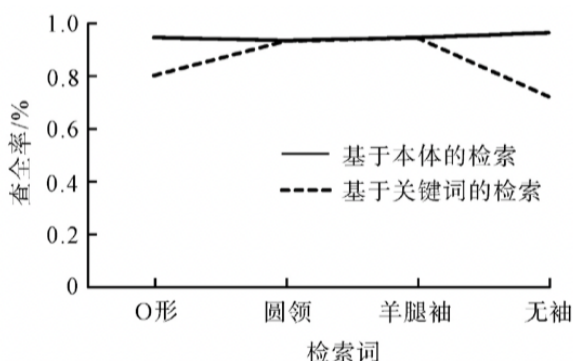


图 3 查全率对比

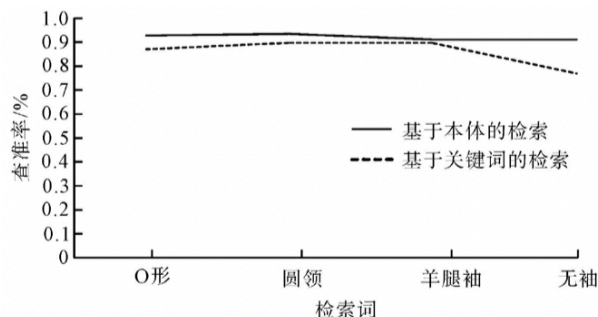


图 4 查准率对比

由表 2、图 3 和图 4 可以看出，若待检索关键词无同义词时，如“圆领”“羊腿袖”，基于本体的检索和基于关键词的检索查全率结果不相上下，而前者查准率较高；但在检索同义词较多的关键词时，如“O 形”“方领”等，基于本体的检索模型明显比基于关键词的查准率和查全率高。因此，基于本体的连衣裙款式检索模型得到了更好的检索效果，在查全率和查准率上得到了明显提高。

6 小结

本文通过建立连衣裙款式领域本体，提出了基于语义层面的连衣裙款式检索模型。通过建立连衣裙款式本体，将连衣裙款式方面的概念进行汇总与整理，结合相似度算法，将用户的检索词进行同义扩展与相似扩展，有效提高了检索的查准率与查全率。与关键词检索方式相比，基于本体的连衣裙款式检索模型能更好地理解用户语义，为用户提供更为精确的检索结果，今后可与图像识别、自动添加标签等技术结合，实现更智能的服装款式检索。

参 考 文 献

- [1] 厉颖. 基于本体的服装领域知识表示的建模研究[D]. 青岛: 山东科技大学, 2010.
- [2] 李永超, 罗钧旻. 语义 Web 中的本体推理研究[J]. 计算机技术与发展, 2007(1): 101-103.
- [3] 李文靖, 胡书山, 余日季. 基于语义网的数字化家具模型本体设计与检索[J]. 软件导刊, 2019, 18(8): 136-139
- [4] WANG Qian. Research on ontology-based garment manufacturing process information context modeling

[J].Software Guide, 2015, 14(12): 118-121.

[5] 张志武.基于本体的 Web 图像语义标注与检索模型[J].情报科学, 2015, 33(4): 121-124.

[6] 梁道雷, 姚亮, 陈巩静.基于本体的纺织服装主题搜索引擎系统研究[J].丝绸, 2014, 51(11): 28-31.

[7] 刘晓刚.服装设计元素论 [J].东华大学学报(自然科学版),2003(2): 23-26.

[8] XAVIERA, SOPHIE G, JEREMY H. Vetu Voc: a modular ontology for the fashion, textile and clothing domain [J].Applied Ontology,2015,11(1):1-28.

[9] 张韩. 连衣裙的造型要素与感性意象关联量化及款式推荐研究[D].杭州: 浙江理工大学, 2017.

[10] 罗仕鉴, 董焯楠. 面向文化创意设计的器物知识集成与管理[J].计算机集成制造系统, 2018, 24(4): 964-977.

[11] ZHANG Naijing, JU Hongbo, JI Ping. Modeling semantic queryexpansion of the forestry domain based on ontology[J].Computer System Application, 2016, 25(3):151-156.