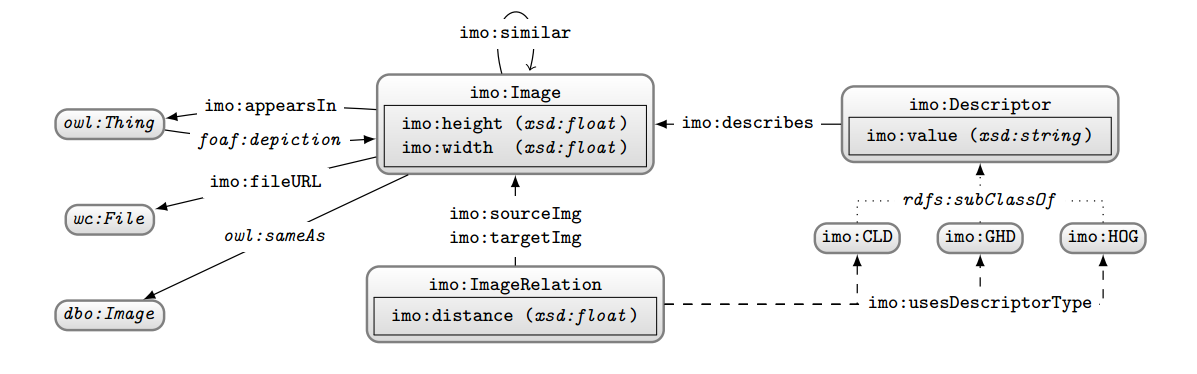
## Richpedia现有不足

1. 文件格式不规范

之前存储关系的文件格式用的是json文件，经过查阅资料将其改为更为规范的nt文件进行存储；同时需要规范一下richpedia的ontology，构建类似于IMGpedia的ontology，并在网页上进行ontology和resource的路由展示。



1. 图像算法过于简单

我们之前的图像算法仅限于一个图像像素直方图的相似计算，通过调研IMGpedia，发现可以通过借助已有的图像描述符来构建图像的特征表示，无论是之后的图像相似度查询还是联合查询都可以利用图像描述符进行辅助计算。

1. 关系发现算法还需增强

之前的关系发现是基于规则的算法，具有较强的约束性，评审意见中也有关于规则有效性的质疑。因为关系发现借助于Wikipedia的超链接和文字描述信息，从而限制了关系的一方至少为Wikipedia中的图像，所以导致richpedia的关系稀疏，需要利用更为有效的关系发现算法。

## 下版本Richpedia的创新点

1. 图像资源来源广阔

对比与IMGpedia和MMKG，IMGpedia主要处理的图像为Wikimedia Commons中存在的图像，图像数量有限，且大部分图像时间较为久远，显得有些陈旧；MMKG主要本体从FB15K等小批量数据集中获取，而且没有建立图像之间的关系。所以Richpedia从互联网中获取新鲜的图像信息，通过加工和梳理，建立起新颖的多模态图像关系。

1. 关系发现定义新颖

在IMGpedia中，图像实体之间并没有建立起多模态关系，只是通过计算图像特征的相似性提供图像间的相似关系，并不是通过上下文语义关系进行发现；MMKG中的关系也是依附于之前存在的关系中，所以也没有建立多模态语义关系。所以构建多种多样的多模态语义关系是Richpedia的一大主要创新点。

## 工作总结

1.文件格式规范化

现已将所有文件（包括从属文件、关系文件和图像特征文件等）规范化为W3C所要求的N-triples格式，Richpedia的Ontology构建工作也基本完成，只需要在网站上建立相应的Ontology界面即可。

2.图像算法改进

现有图像算法的描述符共六个，分别为灰度直方描述符（GHD），方向梯度直方图描述符Sobel算子（HOGS），方向梯度直方图描述符Laplace算子（HOGL），颜色分布直方图（CLD），颜色矩描述符（CM），灰度共生矩阵描述符（GLDM）。

灰度直方图描述符和颜色分布直方图可以描述描述了图像颜色的统计分布特性，两个方向梯度直方图描述符可以构建图像的边缘信息，颜色矩描述符可以在颜色直方图的基础上计算出每个颜色通的均值、方差、偏差，而灰度共生矩阵可通过图像中灰度级分布的随机属性来描述纹理特征，各个特征具体描述见下附。

我们现在拥有的所有图像特征描述符完全包含了IMGpedia所拥有的的特征描述符（IMGpedia为GHD、HOGS和CLD），最终可以在这几种描述符中选取三到四个作为最终的描述符来构建Richpedia的图像特征描述体系，以上所有的描述符均已生成规范的N-triples文件进行数据存储。

1. 关系发现算法改进

发现关系算法主要还是以依据三条规则的算法为主，主要重点为详细叙述算法过程，讲清例如过程中运用的stanford-core nlp工具以及如何利用nlp工具实现借助于超链接和文本信息的过程（主要分为三种情况：含有多个超链接信息，含有一个超链接信息和不含有超链接信息）。

对于上次讨论的达到百科级别的知识图谱，会在之后的过程中实现，因为实现百科级别需要大量的带宽资源和硬件存储容量，而且JIST投稿时间较为紧张，我们先实现上述三点的改进，之后会逐渐将Richpedia的规模加大，以达到百科级别容量。

## 附：详细描述符及算法介绍：

IMGpedia计算的描述符如下：

1. 灰度直方描述符(GHD)：将图像从彩色转换为灰度，并将其划分为固定数量的块。然后计算每个块的8位灰度直方图。所有直方图的连接用于生成一个256维的描述向量。
2. 方向梯度直方图描述符(HOG)：通过计算灰度图像的梯度(使用Sobel算子)，使用一个阈值计算梯度的方向来提取灰度图像的边缘。最后，制作了一个方向梯度直方图，288维的描述向量。

HOG：方向梯度直方图特征是一种在计算机视觉和图像处理中用来物体检测的特征描述子，HOG特征通过计算和统计图像局部区域的梯度方向直方图来构成特征。思想：在一副图像中，局部目标的表象和形状能够被梯度或边缘的方向密度分布很好地描述。其本质为：梯度的统计信息，而梯度主要存在于边缘的地方。

1. 颜色分布直方图（CLD）：将图像分成块，并为每个块计算平均值(YCbCr)颜色。然后对每个颜色通道进行离散余弦变换。最后，将转换的连接用作具有192维的描述符向量。

YCbCr：Y就是所谓的流明(luminance，描述光通量的物理单位)，表示光的浓度且为非线性，而CB和CR则为蓝色和红色的浓度偏移量成份。

CBIR（基于内容的图像检索）：

基于内容的图像检索，即CBIR(Content-based image retrieval)，是计算机视觉领域中关注大规模数字图像内容检索的研究分支。典型的CBIR系统，允许用户输入一张图片，以查找具有相同或相似内容的其他图片。而传统的图像检索是基于文本的，即通过图片的名称、文字信息和索引关系来实现查询功能。

这一概念于1992年由T.Kato提出的。他在论文中构建了一个基于色彩与形状的图像数据库，并提供了一定的检索功能进行实验。此后，基于图像特征提取以实现图像检索的过程以及CBIR这一概念，被广泛应用于各种研究领域，如统计学、模式识别、信号处理和计算机视觉。

特征提取：

底层图像特征包含颜色、纹理、平面空间对应关系、外形，或者其他统计特征。 图像特征的提取与表达是基于内容的图像检索技术的基础。从广义上讲，图像的特征包括基于文本的特征（如关键字、注释等）和视觉特征（如色彩、纹理、形状、对象表面等）两类。视觉特征又可分为通用的视觉特征和领域相关的视觉特征。前者用于描述所有图像共有的特征，与图像的具体类型或内容无关，主要包括色彩、纹理和形状；后者则建立在对所描述图像内容的某些先验知识（或假设）的基础上，与具体的应用紧密有关，例如人的面部特征或指纹特征等。

颜色：

颜色是彩色图像最底层、最直观的物理特征，通常对噪声，图像质量的退化，尺寸、分辨率和方向等的变化具有很强的鲁棒性，是绝大多数基于内容的图像和视频检索的多媒体数据库中使用的特征之一。颜色特征的描述方法主要有以下四种：

1. 颜色直方图(ColorHistogram)

它是最简单也是最常用的颜色特征，描述了图像颜色的统计分布特性，具有平移、尺度、旋转不变性。其核心思想是在颜色空间中采用一定的量化方法对颜色进行量化，然后统计每一个量化通道在整幅图像中所占的比重。

常用的颜色空间有RGB，CIE，HSI，HSV空间等，主要的量化方法有最重要信息位、颜色空间划分、颜色空间聚类、参考颜色、图像分割等，文献中讨论了对这些方法进行了讨论和总结。由于颜色直方图缺乏颜色的空间分布信息，改进的方法包括在颜色索引时加入空间位置信息和基于区域的颜色查询。最简单的方法是子窗口直方图法，即将图像分割成子图像，一一建立索引。另一文献中将图像分成了大小相等的九个子图像，然后统计每个子图像中的颜色直方图。

1. 颜色相关图(ColorCorrelogram)

其主要思想是用颜色对相对于距离的分布来描述信息，它反映了像素对的空间相关性，以及局部像素分布和总体像素分布的相关性，并且容易计算，特征范围小，效果好。

1. 颜色矩(ColorMoment)

其基本思想是在颜色直方图的基础上计算出每个颜色通的均值、方差、偏差，用这些统计量替代颜色的分布来表示颜色特征。它具有特征量少，处理简单的特点。

1. 颜色一致性矢量(Color Coherence Vectors, CCV)

本质上是一种引入空间信息改进的直方图算法，统计了图像中各颜色最大区域的像素数量。通过分离开一致性像素和非一致性像素，比直方图算法具有更好的区别效果。

纹理：

纹理是图像的重要特征之一，通常定义为图像的某种局部性质，或是对局部区域中像素之间关系的一种度量，其本质是刻画像素的邻域灰度空间分布规律。纹理特征描述方法大致可以分为四类：统计法、结构法、模型法、频谱法。

1. 统计法

统计法分析纹理的主要思想是通过图像中灰度级分布的随机属性来描述纹理特征。最简单的统计法是借助于灰度直方图的矩来描述纹理，但这种方法没有利用像素相对位置的空间信息。

为了利用这些信息，Haralick 等人提出了用共生矩阵来表示纹理特征。该方法研究了纹理的空间灰度级相关性，构造出一个基于图像像素间方向和距离的共生矩阵，并且从矩阵中提取出反差、能量、熵、相关等统计量作为特征量表示纹理特征。

Tamura 等人基于人类视觉的心理学研究后提出了一些不同的方法来描述纹理特征，给出了几个不同的描述纹理特征的术语：粗糙度(Coarseness) 、对比度(Contrast) 、方向(Directionality) 、线性度(Linelikeness) 、规则度(Regularity) 、粗略度(Roughness) 等。Tamura 纹理和共生矩阵表示的主要区别在于:前者的所有纹理属性都是视觉意义上的，而后者的某些纹理属性不具有视觉意义(如信息熵) 。这一特点使得Tamura 的纹理表示在图像检索中使用得较多。QBIC 和MARS都进一步证明了这种表示方法。

1. 结构法

结构法分析纹理的基本思想是假定纹理模式由纹理基元以一定的、有规律的形式重复排列组合而成，特征提取就变为确定这些基元并定量分析它们的排列规则。Carlucci曾提出一个使用直线段、开放多边形和封闭多边形作为纹理基元的纹理模型，其排列规则由一种图状语法结构定义。Lu and Fu给过一种树型语法结构表示纹理，他们将纹理按照9 ×9 的窗口进行分割，每个分解单元的空间结构表示为一棵树。因为实际的纹理大都是无规则的，因此结构法受到很大限制。

1. 模型法

模型法利用一些成熟的图像模型来描述纹理，如基于随机场统计学的马尔可夫随机场、子回归模型，以及在此基础上产生的多尺度子回归模型(MultiResolution Simultaneous Autoregressive, MRSA) 等。这些模型的共同特点是通过少量的参数表征纹理。MRSA 区分不同纹理模式的能力较强，但同时计算开销也较大。

1. 频谱法

频谱法借助于频率特性来描述纹理特征，包括傅里叶功率谱法、Gabor 变换、塔式小波变换( Pyramid Wavelet Transform ，PWT)、树式小波变换( Tree Wavelet Transform，TWT)等方法。Manjunath and Ma实验指出， Gabor 特征提供了最佳的模式检索精度，检索性能优于TWT 和PWT，略微优于MRSA ，缺点是计算速度慢，其旋转不变性和尺度不变性仍有待讨论。

形状：

形状是刻画物体最本质的特征，也是最难描述的图像特征之一，主要难在对图像中感兴趣目标的分割。对形状特征的提取主要是寻找一些几何不变量。目前用于图像检索的形状描述方法主要有两类:基于边缘和基于区域的形状方法。前者利用图像的边缘信息，而后者则利用区域内的灰度分布信息。

1. 基于边缘：

基于边缘的形状特征提取是在边缘检测的基础上，用面积、周长、偏心率、角点、链码、兴趣点、傅里叶描述子、矩描述子等特征来描述物体的形状，适用于图像边缘较为清晰、容易获取的图像。文献首先对图像进行了高斯平滑，接着使用经典的兴趣点检测算法发现兴趣点，然后用兴趣点的测度值作为图像特征进行匹配。文献提出将图像边缘上的角点作为特征点，然后使用Delaunay三角形进行划分，记录三角形的形状特征来描述图像的形状特征。这种方法由于是基于边缘上的一些特殊点，因此对噪声和点位置的变化较为敏感。文献采用边缘方向直方图来刻画形状特征，具有简单、平移不变性等优点，但也存在不具备尺度、旋转不变性等缺点。

1. 基于区域：

基于区域的形状特征提取的主要思路是通过图像分割技术提取出图像中感兴趣的物体，依靠区域内像素的颜色分布信息提取图像特征，适合于区域能够较为准确地分割出来、区域内颜色分布较为均匀的图像。文献应用变形模板技术，把用户提供的形状看作模板，与图像库中的形状进行匹配。由于是直接比较两个形状，因此具有较高的精度，但同时计算量也较大。文献提出了一种形状弹性匹配算法，首先确定感兴趣区域，在这些区域中采用爬山优化算法获取图像边缘，并用这些边缘代表物体形状。这种方法的优点是对图像边缘进行了筛选，缺点是需要人工干预。近年来，基于区域的图像检索方法已经成为基于内容的图像检索的一大研究热点。

特征匹配：

从图像中提取的特征可以组成一个向量，两个图像之间可以通过定义一个距离或者相似性的测量度来计算相似程度。

特征匹配是图像检索的一个关键环节，具有特征依赖的特点，不同的特征应该采用不同的度量方法。在检索的过程中，根据系统相似性度量的算法计算查询特征与特征库中对应的每组特征的相似程度，把所得结果由大到小排序后得到一个匹配图像序列返回给用户。其间可以通过人机交互，对检索的结果逐步求精，不断缩小匹配集合的范围，从而定位到目标。匹配过程常利用特征向量之间的距离函数来进行相似性度量,模仿人类的认知过程,近似得到数据库的认知排序。常用的距离度量公式有：Minkkowsky距离，Manhattan距离，Euclidean距离，加权Euclidean距离，Chebyshev距离，Mahalanobis距离等。

其中，Manhattan 距离计算简单，效果也较好，被广泛采用；加权Euclidean 距离考虑了不同分量的重要性,也较为常用；Mahalanobis 距离考虑了样品的统计特性和样品之间的相关性，在聚类分析中经常用到。当采用综合特征进行检索时，需要对各特征向量进行归一化,以使得综合特征的各特征向量在相似距离计算中地位相同。

语义鸿沟：

通常人们在判别图像的相似性时并非建立在图像低层视觉特征的相似上，而是建立在对图像所描述的对象或事件的语义理解的基础上。这种理解无法从图像的视觉特征直接获得，它需要使用人们日常生活中积累的大量经验和知识来进行推理和判断。其中,尤其对于一些高层次的抽象概念，如一幅关于节日的图像所表达出的欢乐和喜庆的感觉等，更需要根据人的知识来判断。换言之，人们是依据图像的语义信息来进行图像相似性判别的。正是由于人对图像相似性的判别依据与计算机对相似性的判别依据之间的不同，造成了人所理解的“语义相似”与计算机理解的“视觉相似”之间的“语义鸿沟”的产生。

在传统的基于文字的查询技术中，不存在这个问题，因为查询关键字基本能够反映查询意图。但是在基于内容的图像查询中，就存在一个底层特征和上层理解之间的差异（这也就是著名的semantic gap）。主要原因是底层特征不能完全反映或者匹配查询意图。弥补这个鸿沟的技术手段主要有：

1.相关反馈（relevance feedback）

按照最初的查询条件，查询系统返回给用户查询结果，用户可以人为介入（或者自动）来选择几个最符合他查询意图的返回结果（正反馈），也可以选择最不符合他查询意图的几个返回结果（负反馈）。这些反馈信息被送入系统用来更新查询条件，重新进行查询。从而让随后的搜索更符合查询者的真实意图。

2.图像分割（image segmentation）

图像的特征可以包括全局特征和局部特征。如果进行一定程度的图像分割，划分出不同的分割区域，这样可以增加局部特征的信息量，也可能在一定程度弥补语义鸿沟。

3.建立复杂的分类模型（Machine Learning）

一些比较复杂的非线性分类模型，比如支持向量机（Support Vector Machine）本身就可以起到一定程度的效果来弥补语义鸿沟。