**面上项目申请书撰写提纲**

**（2018版）**

面上项目申请书由信息表格、正文、个人简历和附件构成。

**一、信息表格：**

包括项目基本信息、项目主要参与者和项目资金预算表，填写时应按操作提示在指定的位置选择或按要求输入正确信息；项目资金预算表应按照《国家自然科学基金资助项目资金管理办法》、《国家自然科学基金项目资金预算表编制说明》认真填写，应保证信息真实、准确。

**二、正文：**参照以下提纲撰写，要求内容翔实、清晰，层次分明，标题突出。**请勿删除或改动下述提纲标题及括号中的文字。**

**（一）立项依据与研究内容**（**建议8000字以下**）：

1．**项目的立项依据**（研究意义、国内外研究现状及发展动态分析，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）；

**研究意义**

随着语义Web和关联开放数据（LOD）研究的快速发展，知识图谱逐渐成为研究热点，然而由于传统的知识抽取大部分都是基于自然语言处理技术对文本进行分析，现有的开放知识图谱或者领域知识图谱的构建，很少考虑将图像作为知识抽取对象，这就使所构建出的知识图谱主要基于文本，在使用知识图谱时仅能够实现单一模态的文本语义查询，无法实现可视化的基于图像语义的查询，这其中一个很重要的原因是由于图像内容的丰富性和人们对其内容理解的主观性。如果能够将多模态知识转换成结构化的知识图谱形式，可以丰富知识图谱的内涵，为领域知识的用户提供多个不同维度的知识，同时还能实现不同模态数据间的跨模态交互检索，为研究者提供更便利的帮助。但要构建多模态的领域知识图谱，还存在以下挑战。

**首先，多模态领域知识抽取结果的准确性问题。**例如，由于图像的本身质量问题或图像内容理解的主观性，导致图像知识抽取的准确性不高，需要通过一定的策略来验证所抽取知识的准确性，进而再与文本知识及其他知识库融合。**其次，多模态知识的语义关联问题。**从不同模态信息中所抽取出的知识虽然语义上是相关的，但可能并不再一个语义空间或抽象层次上，如何把多模态语义进行统一的语义表达，进而进行关联与融合是一个要解决的核心问题，这实际上就是多模态知识融合问题。另外，知识融合过程中由于所抽取出的知识以及领域背景知识库通常都是以三元组形式来表示的，在计算实体语义关联或推理关系时需要利用图算法来实现，存在计算复杂度高、效率低、扩展性差、数据稀疏等问题。**第三，知识融合结果隐藏关联的发现问题。**知识融合过程往往重点考虑的是不同模态知识的关联问题，而在知识融合后形成的多模态知识库中，可能会蕴含更多的隐藏语义关联，如果能发现和找到这些隐藏关联，则可以使多模态知识图谱加强和增值。

针对以上问题，在面向大数据的知识融合技术背景下，本课题拟以金属材料领域为例，探索领域知识图谱构建过程中多模态知识融合的策略与方法。本课题的研究具有较高的学术价值和应用价值：

1）研究文本-图像的多模态知识的评估验证方法，将建立一套合理高效的评估机制，为知识抽取结果度量提供有效的理论依据和计算手段，在进行知识融合之前有效地去除噪声，对提高多模态领域知识图谱的质量具有重要意义。

2）建立基于知识表示学习方法的多模态领域知识融合机制，可以在低维空间中高效的对知识进行计算，有效的解决数据稀疏问题。探索不同抽象层次和语义空间的知识融合机理和方法，可为多模态知识图谱的构建提供理论基础和依据，可以应用于多种不同领域的知识融合，具有重要的理论意义和应用价值。

3）整合特定领域的多模态知识，可以为实现不同模态数据间的跨模态交互检索及可视化查询提供数据基础，可进一步丰富知识图谱的方法和理论，为基于知识图谱的智能应用提供新的实践思路。

**国内外研究现状**

**1）领域知识图谱**

知识图谱(knowledge graph)的概念从语义网技术发展而来，是由谷歌公司首先提出的。目前已经有一些大型开放知识图谱，例如FreeBase，YAGO，DBpedia等。在国内，代表性的知识图谱有搜狗的知立方、百度之心等。就覆盖范围而言，以上知识图谱都属于通用知识图谱，主要应用于智能搜索等领域。通用知识图谱强调融合更多的实体，但准确度不够高，而且包含的专业领域知识相对较少[1]。近年来，在一些领域已经出现了面向领域的知识图谱，例如电影领域的IMDB[2]、音乐领域的MusicBrainz[3]、生物医学领域的BMKN[4]、新闻领域的ECKG[5]、健康领域的SHKG[6]等，然而大部分领域的研究进展还处于起步阶段。以金属材料领域为例，目前已经出现了一些本体知识库，例如Ashino材料本体[7]，Matonto[8]等。

从以上已有的领域知识图谱来看，构建领域知识图谱需要借鉴开放知识图谱的方法及先验知识，同时还要根据领域特点灵活设计方法来降低处理知识的粒度问题。然而总体来说，与开放知识图谱的构建相同，构建领域知识图谱需要解决的一个关键问题就是知识融合，即如何将不同数据源抽取出的知识建立合理的关系。另外，领域知识图谱通常是依靠特定行业的数据来构建，具有特定的行业意义，领域知识图谱的构建是当前知识图谱研究的一个重要方向和趋势。

**2）知识融合**

知识融合和数据融合及信息融合不同，数据融合处理的是原始的数据，信息融合处理的是被加工过的已经建立了一定关联关系的数据，而知识融合的对象是知识。DONG[9]等人指出：“知识融合是将从网络大数据公开的碎片化数据中获取的多源异构、语义多样、动态演化的知识，通过冲突检测和一致性检查，对知识进行正确性判断，去粗取精，将验证正确的知识通过对齐关联、合并计算有机地组织成知识库，提供全面的知识共享的重要方法”。从上述的定义可以看出知识融合主要解决知识质量评估及知识扩充两个关键问题。

**知识质量评估。**由于采用开放域信息抽取技术得到的知识元素有可能存在实体识别错误、关系抽取错误，知识推理得到的知识也不一定正确。另外，各个知识库的质量差异也比较大，不同知识库之间存在大量的知识冲突。引入质量评估的意义在于：可以对知识的可信度进行量化，通过舍弃置信度较低的知识，可以保证知识库的质量[10]。知识质量评估的目的主要就是要建立完善的质量评估技术标准和指标体系。知识质量评估采用的主要方法有[11]：基于贝叶斯估计的方法、基于D-S证据理论、基于模糊逻辑理论的方法、基于概率图模型的方法。以上方法都是考虑了知识获取的不确定性，通过对获取知识进行综合评估，去伪存真，以增加知识的可靠性和置信度，从而提高知识库的实用性。

基于贝叶斯方法和基于D-S证据理论的方法由于要求不同来源知识之间相互独立以及知识先验概率是可预知的等问题，难以有效处理大规模知识的评估。基于模糊集理论的知识评估方法需要凭经验设置知识的模糊规则和隶属函数，较难保证多源异构知识的评估结果稳定性。基于概率图模型的方法是近些年采用较为广泛的知识评估方法。谷歌公司创建的Knowledge Vault[12]知识库，它通过抽取算法自动搜集网上信息，得到一些事实，在知识评估阶段根据指定数据信息的抽取频率对信息的可信度进行评分，进而基于图模型利用可信知识库Freebase的先验知识修正抽取所得的事实，提高了知识图谱中知识的质量。

**知识扩充。**针对知识的组成要素，知识扩充可以分为实体扩充、关系扩充和分类扩充三类[11]。实体扩充的主要目标是将获取的实体动态扩展到知识库中，具体可以分为实体链接和实体分类。关系扩充的主要目标是将获取的实体关系动态扩展到知识库中，采用的主要方法有基于语义的方法和基于嵌入（Embedding）学习的方法。分类扩充的主要目标是将描述知识的两个分类体系进行集成，实现知识的复用和共享，包括分类对齐和分类合并两个部分。

从方法层面分析，知识扩充的方法主要有[13]：基于本体的扩充方法、基于规则的扩充方法、基于统计学习的扩充方法以及基于上下文的扩充方法。

上述的知识评估及知识扩充方法大部分都是基于文本知识的，涉及到针对图像知识的评估与扩充方法相对较少，也就是说目前大部分知识融合的对象往往是单模态的文本知识。而互联网上的图像信息包含了丰富的语义信息，如何评估图像知识并融合到知识库中，该问题已经成为一个不可回避的研究趋势。近些年已经有学者研究集成文本-图像多模态知识库的构建，具体内容详见下文的综述。

**3）文本-图像多模态知识融合**

多模态知识融合可以分为两个层面，一个是特征层面的融合，目前通常采用多模态深度学习方法。另外一个是知识层面的融合，可以通过知识表示学习方法实现[14-16]。

多模态深度学习。深度学习理论的快速发展，极大的推动了自然语言理解、信息检索、计算机视觉等众多领域的进步，而多模态深度学习为解决文本-图像的特征融合提供了一个有效的途径。基于文本和图像的多模态深度学习旨在将文本和图像通过深度学习方法进行语义融合，从而为图像添加标签或者根据文本描述生成图像。随着Ngiam[17]等人提出多模态深度学习方法，计算机视觉领域开展了大量的诸如图像标题生成[18]，文本条件图像生成[19]等多模态语义融合相关的研究工作。Vendrov[20]等人认为图像和文本在更高层的语义空间存在关联，并依此进行图像和文本的语义融合处理，从而实现通过自然语言对图像语义的检索。近年来，研究人员还尝试把多模态深度学习方法应用到诸如视觉问答（Visual Question Answering, VQA）、多模态信息检索等领域的研究工作[21-23]。多模态深度学习本质上是通过学习得到多模态数据的高层特征联合表示，从而能够较好的实现多模态语义融合，但在该过程中并没有形成可显性表达的知识表示，而在领域知识图谱构建工作中的知识融合对象往往是显性知识，因而直接应用多模态深度学习方法还存在特征层面到知识层面映射的问题。

知识表示学习。表示学习又称为特征表示学习，主要研究如何更好的学习到原始数据的高层特征表示，从而实现不同模态信息的高层语义融合。知识表示学习是将研究对象的语义信息表示为稠密低维实值向量，可以在低维空间中高效计算实体和关系的语义联系，有效解决数据稀疏问题，使知识获取、融合和推理性能得到显著提升[24]。Bordes[25]等人提出了面向知识图谱的分布式表示学习方法TransE，使得在知识图谱相关的研究中不必再采用传统的字符串匹配的算法，而是在低维的向量空间中进行关系预测等知识计算工作。基于文献[17]的工作，Zhang[26]等人提出了可视化关系检测模型VTransE，该模型可以应用到视觉问答领域。知识表示是知识获取、融合及应用的基础，知识表示学习问题是贯穿知识库的构建与应用全过程的关键问题。知识表示学习的出现为图像知识与文本知识的高层融合提供了有效的方法，但是传统的知识表示学习都是以知识库中三元组形式存储的知识为训练数据，而多源信息融合能够有效提升知识表示的性能[16, 27, 28]。在文本-图像多模态知识融合过程中，同时利用知识三元组、图像及周边上下文等多源信息来共同作为学习目标，从而得到一个表示能力更强的知识表示模型，将能更好的提升多模态知识融合的效果。

多模态知识库方面。为了构建集成可视化知识的多模态知识库，Chen[29]等构建了NEIL系统，通过半监督学习方式从互联网上的大量图像资源中抽取可视化知识。DONG X[12] 等尝试在知识库中集成可视化的图像知识，但仅仅是通过超级链接的方式连接到文本。SADEGHI[30]等试图从图像中抽取知识,并建立一个可视化知识抽取系统，通过可视化的方式来验证文本知识中实体之间关系短语的正确性。知识库的结构上是由节点和边构成的图结构，其中节点表示实体，而边则表示实体之间的关系。图像中往往包含丰富的语义信息，从知识库的角度看，每个图像都可以看作是一个独立实体或者实体的集合，在计算机视觉领域中常使用马尔科夫随机场（Markov Random Filed，MRF）这种概率图模型来完成相关工作。ZHU[31]等就是采用在知识库表达中引入大规模MRF模型的方法，将图像与文本及结构化数据融合到一起，从而实现了一个能够进行灵活高效的可视化查询的多模态知识库。Steffend[32]等人提出了一种基于文本、图像以及已有知识库集成的多模态知识库构建方法。

综合以上对多模态知识融合方法及多模态知识库构建等方面现有研究工作的分析，可以看出，融合多模态知识并集成到知识库中，从而实现一个可视化的知识图谱是非常有意义的热点研究方向，是当前的研究趋势。针对领域知识图谱构建中的多模态知识融合，如何利用知识表示学习方法来学习一个表示能力更强的高性能模型是关键问题，而这其中的核心科学问题是如何对图像及周边上下文进行编码并集成到三元组中来共同作为学习目标。上述文献中，对领域知识图谱构建过程中该问题的描述还较为少见。

综合以上三部分对国内外研究现状的分析，不难发现在构建领域知识图谱过程中考虑多模态知识融合是当前发展趋势下的重要问题。本课题拟以金属材料领域为例，探索领域知识图谱构建过程中文本-图像多模态知识评估和知识融合的方法。本课题在完成领域知识融合时拟引入知识表示学习的方法，实现多模态知识融合。本课题的研究将为多模态知识图谱的构建提供理论基础和依据，进一步丰富知识图谱的方法和理论。 课题组已经在面向YAGO的金属材料领域本体构建[33]、面向DBpedia的金属材料领域本体构建[34]、基于映射结果的本体扩展查询[35]领域进行了系统的研究，为课题的开展奠定了良好的前期工作基础。

**主要参考文献目录**

[1] 徐增林, 盛泳潘, 贺丽荣, 王雅芳. 知识图谱技术综述 [J]. 电子科技大学学报, 2016, 45(4): 589-606.

[2] IMDB Official . IMDB[EB/OL]. http://www.imda.com.

[3] MetaBrainz Foundation. Musicbrainz [EB/OL]. http://musicbrainz.org/.

[4] TIAN B, GONG L, WANG Y, WANG Y, KULIKOWSKI C A, HUANG L. A method for exploring implicit concept relatedness in biomedical knowledge network [J]. Bmc Bioinformatics, 2016, 17(9): 265-278.

[5] ROSPOCHER M, ERP M V, VOSSEN P, FOKKENS A, ALDABE I, RIGAU G, SOROA A, PLOEGER T, BOGAARD T. Building event-centric knowledge graphs from news [J]. Web Semantics Science Services & Agents on the World Wide Web, 2016, s 37–38(1): 132-151.

[6] SHI L, LI S, YANG X, QI J, PAN G, ZHOU B. Semantic Health Knowledge Graph: Semantic Integration of Heterogeneous Medical Knowledge and Services [J]. BioMed Research International, 2017, 1(4): 1-12.

[7] ASHINO T. Materials Ontology: An Infrastructure for Exchanging Materials Information and Knowledge [J]. Data Science Journal, 2010, 9(9): 54-61.

[8] CHEUNG K, HUNTER J, DRENNAN J. MatSeek: An Ontology-Based Federated Search Interface for Materials Scientists [M]. IEEE Educational Activities Department, 2009.

[9] DONG X L, GABRILOVICH E, HEITZ G, HORN W, MURPHY K, SUN S, ZHANG W. From data fusion to knowledge fusion [J]. Proceedings of the Vldb Endowment, 2014, 7(10): 881-892.

[10] 刘峤, 李杨, 段宏, 刘瑶, 秦志光. 知识图谱构建技术综述 [J]. 计算机研究与发展, 2016, 53(3): 582-600.

[11] 林海伦, 王元卓, 贾岩涛, 张鹏, 王伟平. 面向网络大数据的知识融合方法综述 [J]. 计算机学报, 2017, (1): 1-27.

[12] DONG X, GABRILOVICH E, HEITZ G, HORN W, LAO N, MURPHY K, STROHMANN T, SUN S, ZHANG W. Knowledge vault: a web-scale approach to probabilistic knowledge fusion [C]// proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2014. 601-610.

[13] XUELU Y U, QIAO L. Knowledge Fusion Methods: A Survey [C]// proceedings of the International Conference on Software, Multimedia and Communication Engineering,2017. 286-290.

[14] BENGIO Y, COURVILLE A, VINCENT P. Representation Learning: A Review and New Perspectives [J]. IEEE Trans Pattern Anal Mach Intell, 2013, 35(8): 1798-1828.

[15] CHEN X, DUAN Y, HOUTHOOFT R, SCHULMAN J, SUTSKEVER I, ABBEEL P. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets [C]// proceedings of the Advances in Neural Information Processing Systems,2016. 2172-2180.

[16] XIE R, LIU Z, JIA J, LUAN H, SUN M. Representation Learning of Knowledge Graphs with Entity Descriptions [C]// proceedings of the AAAI,2016. 2659-2665.

[17] NGIAM J, KHOSLA A, KIM M, NAM J, LEE H, NG A Y. Multimodal Deep Learning [C]// proceedings of the International Conference on Machine Learning, ICML 2011, Bellevue, Washington, Usa, June 28 - July,2011. 689-696.

[18] VINYALS O, TOSHEV A, BENGIO S, ERHAN D. Show and tell: A neural image caption generator [C]// proceedings of the Computer Vision and Pattern Recognition,2015. 3156-3164.

[19] REED S, AKATA Z, YAN X, LOGESWARAN L, SCHIELE B, LEE H. Generative Adversarial Text to Image Synthesis [C]// proceedings of the International Conference on Machine Learning, New York, NY, USA,2016. 1060-1069.

[20] VENDROV I, KIROS R, FIDLER S, URTASUN R. Order-Embeddings of Images and Language [J]. arXiv preprint arXiv:151106361, 2015.

[21] FUKUI A, DONG H P, YANG D, ROHRBACH A, DARRELL T, ROHRBACH M. Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding [J]. arXiv preprint arXiv:160601847, 2016.

[22] RAYMOND C, GRAVIER G. Multimodal and Crossmodal Representation Learning from Textual and Visual Features with Bidirectional Deep Neural Networks for Video Hyperlinking [C]// proceedings of the ACM Workshop on Vision and Language Integration Meets Multimedia Fusion,2016. 37-44.

[23] WANG K, HE R, WANG L, WANG W, TAN T. Joint Feature Selection and Subspace Learning for Cross-Modal Retrieval [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(10): 2010-2023.

[24] 刘知远, 孙茂松, 林衍凯, 谢若冰. 知识表示学习研究进展 [J]. 计算机研究与发展, 2016, 53(2): 247-261.

[25] BORDES A, USUNIER N, WESTON J, YAKHNENKO O. Translating embeddings for modeling multi-relational data [C]// proceedings of the International Conference on Neural Information Processing Systems,2013. 2787-2795.

[26] ZHANG H, KYAW Z, CHANG S F, CHUA T S. Visual Translation Embedding Network for Visual Relation Detection [C]// proceedings of the CVPR,2017. 3107-3115.

[27] XIE R, LIU Z, CHUA T, LUAN H, SUN M. Image-embodied Knowledge Representation Learning [J]. international joint conference on artificial intelligence, 2017, 3140-3146.

[28] XIE R, LIU Z, SUN M. Representation learning of knowledge graphs with hierarchical types [C]// proceedings of the International Joint Conference on Artificial Intelligence,2016. 2965-2971.

[29] CHEN X, SHRIVASTAVA A, GUPTA A. NEIL: Extracting Visual Knowledge from Web Data [C]// proceedings of the IEEE International Conference on Computer Vision,2013. 1409-1416.

[30] SADEGHI F, DIVVALA S K, FARHADI A. VisKE: Visual knowledge extraction and question answering by visual verification of relation phrases [C]// proceedings of the Computer Vision and Pattern Recognition,2015. 1456-1464.

[31] ZHU Y, ZHANG C, R C, FEI-FEI L. Building a Large-scale Multimodal Knowledge Base System for Answering Visual Queries [J]. arXiv preprint arXiv:150705670, 2015.

[32] THOMA S, RETTINGER A, BOTH F. Knowledge Fusion via Embeddings from Text, Knowledge Graphs, and Images [J]. arXiv preprint arXiv:170406084, 2017.

[33] ZHANG X, PAN D, ZHAO C, LI K. MMOY: Towards deriving a metallic materials ontology from Yago [J]. Advanced Engineering Informatics, 2016, 30(4): 687-702.

[34] ZHANG X, LIU X, LI X, PAN D. MMKG: An approach to generate metallic materials knowledge graph based on DBpedia and Wikipedia [J]. Computer Physics Communications, 2016, 211(February): 98-112.

[35] ZHANG X, LI K, PAN D. Concepts Expansion in Materials Domain Based on Ontology Matching Results [C]// proceedings of the International Conference on Semantics, Knowledge and Grids,2017. 141-145.

2．**项目的研究内容、研究目标，以及拟解决的关键科学问题**（此部分为重点阐述内容）**；**

**研究内容**

**（1）基于上下文及背景知识库的多模态知识评估验证方法**

针对从多模态信息源抽取的知识，构建知识评估验证模型和验证方法，用于验证所抽取知识中的实体及实体关系的准确性和可信性。**验证模型**包括知识验证的各项指标设计、以及对应的知识验证算法，作为知识验证的尺度标准和度量依据。**验证方法**分为两个层次：①通过建立基于背景知识库的先验模型，验证从文本信息源抽取的知识可信度，并赋予相应的置信度值；②通过基于图像标题、周边文本等建立的上下文模型，来验证视觉技术所抽取的图像语义的可信度，同样也赋予相应的置信度值。

**（2）基于表示学习和知识置信度的多模态领域知识融合**

本研究以金属材料领域为例，构建用于表征多模态领域知识的表示学习方法，建立知识表示学习模型，通过该模型将文本知识与图像知识映射到统一的低维语义空间，实现领域多模态知识的融合。为了更准确的表达领域知识，提高领域知识表示的性能，建立多源多模态的知识表示学习机制，在表示学习过程中考虑知识库中实体的描述文本和分类信息，以及实体对应的图像信息等多源多模态信息。由于知识获取的不确定性，试图在知识融合过程中建立基于知识置信度的权重机制，以提高知识图谱中知识的质量。

**（3）基于语义关联的多模态知识融合结果扩展方法**

在多模态知识库的基础上，探索图像语义与开放知识库的关联机制，建立基于图像上下文信息及图像视觉特征的语义关联扩展方法。具体研究以下两方面的内容：①利用图像上下文语义模型及图像视觉特征作为图像语义的扩展，实现图像语义与开放知识库关联，增加图像的相关语义；②利用增强后的图像相关语义，探索图像之间隐含语义关联的计算方法，实现图像关联的强化和补全。

**研究目标（定量指标和定性指标相结合）**

本课题的总体目标是研究构建领域知识图谱过程中文本-图像多模态知识融合的方法。课题将以金属材料领域为例，探索评估验证所抽取文本-图像多模态知识正确性的方法；研究联合多源、多模态信息的知识表示学习方法，以此来提高知识表示性能，从而建立更有效的文本-图像多模态表示学习模型；研究基于表示学习模型的图像与文本多模态知识融合方法，构建二者向统一语义空间中的映射机制，实现图像语义与文本语义在低维、稠密向量空间中的融合；通过建立图像资源与开放知识库的关联的方式，探索知识图谱中的隐含关系，进一步扩展知识融合的结果。该研究的成功将为多模态知识评估与知识融合领域提供理论和实践依据。

**拟解决的科学问题**

（1）面向领域的图像知识正确性评估机理。面向特定的领域，利用视觉技术从图像中所抽取语义的准确性通常较低。因此，利用图像周边上下文和背景知识库来对图像知识的正确性进行评估，是文本图像多模态知识进行融合的前提和基础，是本课题的重要科学问题。

（2）领域多模态知识置信度度量及不确定性建模机制。通过信息抽取技术所抽取的文本图像多模态知识是具有不确定性的。因而，度量多模态知识的置信度，并建立模型来表达不确定性知识间的复杂关系，是提高知识图谱中知识质量的关键环节。

（3）领域多模态知识融合机制。文本知识和图像知识之间由于语义空间不同或抽象层次不同，存在语义鸿沟。所以，如何把多模态知识通过合理的表示模型来映射到同一个语义空间中，建立统一的表示空间，是实现多模态知识融合的关键问题。同时表示模型的建立为知识图谱中实体的语义相似度计算和知识图谱后续的补全工作奠定了基础。

3．**拟采取的研究方案及可行性分析**（包括研究方法、技术路线、实验手段、关键技术等说明）；

**研究方法**

本课题将采取继承、学习、创新的指导思想，总体研究方案按照“现状分析——>理论研究——>模型设计——>算法设计——>实验原型实现——>评价与优化——>应用实例研究”的方法展开研究。

以下部分按照上文中所提出的研究内容和研究目标来详细阐述课题的技术路线、实验手段及所采用的关键技术。整体技术路线如图1所示。



图1 技术路线图

**基于上下文及背景知识库的多模态知识评估验证方法**

本课题通过在验证模型中设计各项指标及验证算法来初步验证知识的可信度。指标主要从若干个不同的维度来度量所抽取知识中实体与关系的固有及统计特征。本课题拟通过以下几个维度来设计指标：知识来源（文本、图像）、出现频度、内部关联聚合度、一致性等。验证算法主要基于上述单维度或多维度指标来对知识进行匹配验证及合理性的预估。在此基础上，本课题拟通过基于上下文及背景知识库的验证方法来进一步度量知识的置信度值。

对象的语义通常是由其内容和上下文所定义，“上下文”是指可能影响对象语义的所有相关信息和环绕条件。在一个包含领域知识的Web页面环境下，上下文环境包含了图像、图像标题、图像周边文本等。利用视觉技术所获取的图像知识需要解决低层视觉特征与图像的高层语义间的语义鸿沟，该过程中会出现图像知识偏差的问题。利用图像所伴随着上下文文本描述信息作为图像的“语义”特征，可以辅助判断图像知识的正确性。反之，文本中也存在大量的无关信息，导致所抽取的文本知识中存在噪声。因而在不同数据集中找出实体及关系的描述记录，主要目的是对不同数据源中的实体及关系信息进行整合，形成更加全面的信息，是评估过程中基本问题，判断多模态的知识是否可信。统计关系学习（Statistical Relational Learning，SRL），集数据的一阶逻辑或关系表示、不确定性建模和机器学习于一体，主要目的是挖掘关系数据中的似然模型。借助背景知识库通过统计关系学习可以获得SRL模型，通过该模型预测所抽取的多模态知识的正确性及其不确定性程度。本课题利用上下文模型以及基于背景知识库的统计关系学习来对多模态知识评估验证，具体技术流程如图2所示。



图2 多模态知识评估验证

（1）为获得可训练的上下文模型，需要确定如何选择分析目标的上下文对象集。根据Markov随机场论(MRF)，分析目标的任何近邻系统都可以建模为该分析目标的上下文。以图像语义作为分析目标，设定其上下文对象集合包含图像标题、图像分类信息、图像周边描述文本等。//各个步骤实例化，尽量贴近咱们的词语。

（2）建立以图像语义为中心的上下文模型。假设图像语义和其上下文对象之间存在的依赖关系，我们可以用某种形式的概率依赖关系(统计相关或统计独立性)来描述上下文依赖关系。提供一种可训练的模型来利用上下文信息对所抽取的知识进行验证，并能进行相应的上下文概率的计算。

（3）上下文模型训练。给定训练数据，可以基于最大似然估计或最大后验估计原则，通过最大期望算法进行模型训练，最终得到描述上下文概率依赖关系的上下文模型。

（4）学习SRL模型。基于背景知识库来生成训练集，以背景知识库中存在的三元组作为正例样本集，而负例样本集拟基于LCWA方法来生成。基于生成的训练集，通过统计关系学习（SRL）方法获得先验SRL模型。

（5）结合利用上下文模型和SRL模型两个工具，通过一定的权重机制来评估多模态知识的置信度，并最终给每个三元组知识赋予相应置信度值。其中权重机制可以作为模型工具的辅助策略，用于灵活调节基于模型所生成的知识置信度值的大小，补偿模型可能产生的偏差。

**基于知识表示学习和知识置信度的多模态知识融合**

多模态知识融合建立在知识获取的基础上，主要包含多模态实体链接和知识合并两部分。本研究将拟采用DBpedia Spotlight提供的策略，同时将YAGO及金属材料领域已有的本体知识库等作为背景知识库的扩展，完成实体的链接。在知识合并方面，从数据层融合出发，将从Wikipedia抽取的知识填充到现有的知识库中，在完成知识合并时拟引入知识表示学习的方法，将通过知识评估验证环节所获得的带有置信度值的知识，利用所学得的知识表示学习模型映射到统一的语义空间中，实现多模态知识融合。具体步骤如下：

（1）将实体回标到文本语料及图像源中；

（2）以此获取到实体词与其他重要单词的共现网络，该网络可以看作联系知识图谱与文本、图像信息的纽带；

（3）基于此网络，定义实体与关系的文本上下文，并将其融入到知识图谱中；

（4）采用TransE模型对实体与关系表示进行学习，获得知识表示模型，利用该模型将达到预设置信度要求的多模态知识向统一的低维语义空间映射，实现融合。

为了提高知识表示的精度，表示学习采用联合文本、图像多模态信息的知识表示学习方法。借鉴清华大学刘知远等提出的文本提升表示学习方法，本课题在考虑文本的同时，还将在训练模型时融入图像知识。具体的方法是通过将知识库中实体所关联的文本与图像进行混合编码的方式来实现。多模态信息编码如图3所示，其中图像编码拟采用AlexNet多层神经网络，文本编码拟采用CNN卷积网络。通过联合映射矩阵将二者映射到实体向量空间，形成多模态信息编码向量。



图3 图像文本多模态编码//注意字体大小

本课题所建立的领域多模态知识表示学习模型如图4所示。

**hs**表示知识三元组的首部实体向量，**hI+T**表示与首部实体关联的多模态信息编码向量，**ts**表示知识三元组的尾部实体向量，**tI+T**表示与尾部实体关联的多模态信息编码向量。



图4 领域多模态知识表示学习模型

**基于语义关联的多模态知识融合结果扩展方法**

本课题将扩展方法划分为三个层次。第一个层次是指如何提取金属材料领域相关的图像信息及元数据的问题，第二层次是如何最大限度的挖掘图像关联的语义信息问题，第三层次是如何处理图像之间的关系以及如何应用问题。前两个层次的过程如图5所示。



图5 图像语义扩展及互联

由于Wikimedia Commons在设计时并没有对数据进行严格的领域划分，因此我们需要设计合理的方法将金属材料领域相关的图像信息提取出来，这部分工作可以视为整个过程的基础。从Wikimedia Commons中获取图片描述、标题、子标题及分类信息等，并从DBpedia Commons中获取当前图像的元数据，包括图像的格式、作者、日期等。其中图像的类别信息及文本描述是挖掘的主要内容，目标是挖掘实体信息并将其链接到DBpedia资源。对图像视觉特征处理的预期目标，则是利用对图像特征的语义标注建立图像与WordNet资源之间的关联。为了解决第三个层次的问题，拟采用相似度计算方法（例如Jaccard），利用图像关联的来自开放知识图谱的资源度量图像之间的语义相似程度，并在基础上，进一步构建查询系统。具体过程如图6所示。



图6 基于互联语义的图像相似计算

图６中LCW表示从Wordnet中获取的图像A的直接关联概念集合，LCD表示从DBpedia中获取的图像A的直接关联概念集合。具体步骤如下：

第一步：图像信息及元数据提取。这一步主要利用Wikimedia Commons提供的查询接口获取金属材料概念相关的图像信息，包括图像的名称、分类信息及描述信息、标题信息等，并获取图像在DBpedia Commons中整理的图像元数据，包括作者、格式、大小等。

第二步：生成图像的直接关联概念集合LCD。这一步主要对图像关联文本进行处理。利用DBpedia对图像关联文本中的实体概念进行注释，将其视为与图像存直接关联的实体概念，生成当前图像的直接关联概念集合LCD。

第三步：生成LCW。我们利用ImageNet训练深度神经网络模型，将其用于提取图像的视觉特征，对特征进行分类并添加语义标签，获取标签在Wordnet中的资源标识，将其视为与图像直接关联的实体概念，生成生成当前图像的直接关联概念集合LCW。

第四步：计算图像相似度。利用图像关联的来自开放知识图谱的资源生成每个图像的扩展概念集合，利用加权Jaccard相似度系数计算图像之间的相似度。

第五步：构建数据集。构建轻量级的本体，保存图像的基本信息、图像与DBpedia资源的关联关系以及图像之间的相似度关系，进一步构建可视化查询接口。

**软件工具原型系统构建及方法验证**

本课题的原型系统构建的具体实施步骤分为数据处理、多模态知识抽取与融合及知识扩展三个步骤。① 数据处理。主要工作是采用统计机器学习方法构建分类器构建领域文本分类器，该分类器用于判断输入的信息是否为领域信息源。采用正确率、召回率、F1值对分类器性能进行评估。②多模态知识抽取。图像知识抽取采用现有的视觉表示学习方法（拟采用深度学习模型VGG-NET）来对图像内容中的实体进行识别。文本知识抽取主要完成从领域文本中自动抽取指定类型的实体及实体关系，形成结构化的数据输出。在DBpedia Spotlight基础上，扩展背景知识库，实现特定领域文本的实体提取。在实体关系抽取方面，本研究利用Wikipedia及DBpedia中已经存在的大量结构化领域知识进行关系类型抽取，并尝试通过建立自适应的关系抽取算法来获取实体关系值。③多模态知识融合与扩展。利用知识表示模型实现多模态知识融合，把融合后的领域知识通过语义扩展方法，将其与现有的开放知识库进行互联，并实现知识库内部图像语义关联的补全与强化。

基于java。。。设计xx软件支撑工具原型系统，具有xxxx1,2，3功能。

基于百科类xxx抽取。。。文本，图像（vggnet）数据；效果验证

4．**本项目的特色与创新之处；**

（1）通过信息抽取技术所抽取的文本图像多模态知识是具有不确定性的，因而，本课题探索结合上下文模型和基于背景知识库的先验SRL模型来进行知识不确定性建模，表达不确定性知识间的复杂关系，并给出知识的置信度，以提高知识图谱中知识的质量。

（2）基于知识表示学习模型的多模态知识融合机制。图像知识和文本知识需要通过合理的知识表示学习模型来映射到同一个语义空间中，建立统一的表示空间，实现语义融合，融合过程中还将考虑知识的置信度水平。表示学习模型的建立为知识图谱中实体的语义相似度计算和知识图谱后续的补全工作奠定了基础。

5．**年度研究计划及预期研究结果**（包括拟组织的重要学术交流活动、国际合作与交流计划等）。

年度研究计划

（1）2019年1月—12月：文献调研；面向从多模态信息源抽取的知识，构建知识评估验证模型和验证方法；研究知识置信度度量机制，构建多模态知识不确定性模型；本年度拟进行国内调研3人次，参加在国内举行的国际/国内学术会议3人次。

（2）2020年1月—12月：构建用于表征多模态领域知识的表示学习方法，建立知识表示学习模型，在知识融合过程中建立基于知识置信度的权重机制；本年度拟进行国内调研3人次，邀请专家交流1次，参加在国内举行的国际/国内学术会议3人次。

（3）2021年1月—12月：基于语义关联的多模态知识融合结果补全和扩展方法研究；本年度拟进行国内调研2人次，国内调研2人次，参加在国内举行的国际/国内学术会议3人次。

（4）2022年1月—12月：设计面向领域知识图谱的多模态知识评估验证和知识融合软件工具原型系统。本年度拟参加在国内举行的国际/国内学术会议3人次；汇总研究成果和各阶段技术报告，完成课题研究报告，进行课题验收。

预期研究结果

（1）理论与方法研究：面向基于多源多模态领域知识图谱的构建，提供有效的知识验证模型和方法，提出一套基于表示学习的知识融合机制和方法。

（2）实验系统：依据所提出的研究方法和策略，构建知识融合工具原型系统。

（3）学术论文：在国内外学术期刊和国际学术会议上发表学术论文12篇以上，且被SCI或EI收录。

（4）人才培养：培养硕士研究生4-6名。

**（二）研究基础与工作条件**

1．**研究基础**（与本项目相关的研究工作积累和已取得的研究工作成绩）；

本课题组一直致力于科学数据管理、机器学习、语义计算等方面的研究工作，在语义Web、信息集成、知识图谱等领域具有较好的研究基础，形成了良好的合作机制。

课题申请人张晓明在2013年至2017年期间，主持或参研与本项目相关的课题多项，具体所做的工作积累如下。

（1）国家自然科学基金面上项目“金属材料科学数据语义集成关键技术研究”（51271033，2013-2016）。与北京科技大学合作承担，河北科技大学方负责人。负责金属材料知识语义表示模型的构建、语义映射研究、语义查询研究。在知识的领域识别、半结构化数据抽取领域知识、基于机器学习的领域本体实例填充等方面设计了具有领域特色的方法和算法，同时在金属材料领域积累了大量的实验数据和领域知识。

（2）河北省自然科学基金“云计算环境下的异构传感网数据语义集成关键技术研究”（F2013208107，2013-2016），主持人。通过该项目的研究，在异构传感网数据的知识抽取、表示并形成知识图谱方面做了探索和尝试，为知识融合方面的研究奠定了良好基础。

（3）国家自然科学基金面上项目“基于语义距离的分布式数据挖掘理论与方法”（71271076，2013-2016），排名第二. 负责语义距离度量体系。通过该项目的研究，在语义相关性和语义相似性度量方面，设计并应用了多种度量策略和算法，为知识的评估验证方面积累了相关经验。

课题申请人近年来发表知识图谱、材料信息学方面的论文20余篇（其中第一作者SCI收录论文5篇），形成了扎实的研究基础：（1）从DBpedia和YAGO通用知识图谱识别领域知识的方法和策略；（2）从Web数据表格抽取领域信息形成领域知识图谱的方法和策略（3）基于Wikipedia的金属材料图片知识抽取方法；（4）金属材料本体的构建与知识的半自动化填充方法等。

通过以上工作既为本课题的研究积累了基础方法和实践经验，也培养了申请人主持科研项目的基本能力。此外申请人具有材料信息学的研究背景，为本研究积累了一定数量的数据基础，能够支持本课题正式展开后的前期工作，为本课题的实证研究奠定基础。申请者有承担本课题研究的学术梯队、工作经验和组织能力。

课题组其他主要研究人员近年来参研国家级、省级课题多项，获得软件著作权登记10余项，在数据智能处理、语义计算、知识图谱和机器学习等领域具有较好的研究基础。其中高凯老师作为主要成员参加国家自然科学基金面上项目“语义主题与社交关系融合的特定群体发现关键技术研究”（61272362，2018-2021，与北京理工大学合作承担，河北科技大学方负责人，总排名第二），负责基于深度学习方法完成面向融合语义主题与社交关系的群体发现算法；主持河北省自然科学基金项目“基于微博文本情感与博主剖像的用户行为分析研究”（F2017208012， 2017-2019），负责基于机器学习与Seq-2-Seq、GAN等深度学习算法的文本挖掘与生成工作。这些工作为机器学习算法的应用积累了丰富的研究经验，可以为本课题提供很好的研究基础。

王会勇老师作为主研人参与国家自然科学基金项目“面向并行程序中同步机制的软件自动重构及一致性检验方法研究”（61440012，2015-2016），负责软件重构算法的实现；参与石家庄市科技计划项目“基于物联网技术的机房智能监控系统”（16SCX01004，2016-2017），负责视频监控、温度、湿度等多模态传感器数据的融合算法设计工作；主持与国有大中型企业合作研发项目“解调器监控模块编码软件”，负责解调器数据采集、融合及可视化方面的工作。王会勇老师具有数据分析、信息融合、系统开发方面研究经验，这些可以为本课题中图像知识获取提供很好的研究基础，也为本课题的实证研究积累了设计和开发经验。

课题组成员分别来自语义Web、数据挖掘、机器学习和材料信息学等研究方向，组成了一支方向交叉的科研团队，通过专题讨论和课题预研，这支团队已经能从多角度、多层次分析和求解问题，有助在本课题关键环节的研究中产生具有特色和创新性的成果。

2．**工作条件**（包括已具备的实验条件，尚缺少的实验条件和拟解决的途径，包括利用国家实验室、国家重点实验室和部门重点实验室等研究基地的计划与落实情况）；

申请者所在单位河北科技大学信息科学与工程学院计算机系的“计算机软件与理论”是**河北省省级重点学科，**学院能够提供本项目研究所需的实验场所，计算机系可以为本课题的研究提供有以下设备条件：

（1）戴尔PowerEdge R710（Intel Xeon 5620\*2，12G 内存, 4\*146G）机架式服务器4台；戴尔PowerEdge 1950 （Inter Xeon E5430） 机架式服务器 8台。

（2）戴尔NX3100网络存储（Intel Xeon 5620\*2, 6TB）；群晖DS916+ NAS网络存储服务器，16TB 。

（3）戴尔Precision T5810图形工作站4台（E5-1620 v3 CPU，32G 2133MHz DDR4，W5100显卡4G）。戴尔移动工作站2台。

（4）高性能GPU服务器1台。

本课题的研究不需要再购置大型设备和仪器，所需的实验条件已经具备。

3．**正在承担的与本项目相关的科研项目情况**（申请人和项目组主要参与者正在承担的与本项目相关的科研项目情况，包括国家自然科学基金的项目和国家其他科技计划项目，要注明项目的名称和编号、经费来源、起止年月、与本项目的关系及负责的内容等）；

无

4．**完成国家自然科学基金项目情况**（对申请人负责的前一个已结题科学基金项目（项目名称及批准号）完成情况、后续研究进展及与本申请项目的关系加以详细说明。另附该已结题项目研究工作总结摘要（限500字）和相关成果的详细目录）。

无

**（三）其他需要说明的问题**

1. 申请人同年申请不同类型的国家自然科学基金项目情况（列明同年申请的其他项目的项目类型、项目名称信息，并说明与本项目之间的区别与联系。

无

2. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在同年申请或者参与申请国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，申请或参与申请的其他项目的项目类型、项目名称、单位名称、上述人员在该项目中是申请人还是参与者，并说明单位不一致原因。

无

3. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在与正在承担的国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，正在承担项目的批准号、项目类型、项目名称、单位名称、起止年月，并说明单位不一致原因。

无

4. 其他。

无

**三、个人简历**：

1. 申请人简历（由系统根据申请人在线填写的个人简介信息、承担项目情况和个人研究成果自动生成）

2. 主要参与者简历（在读研究生除外）（请下载参与者简历模板填写后上传；**除非特殊说明，请勿删除或改动简历模板中蓝色字体的标题及相应说明文字**）

**四、附件**

**（一）附件目录**

在附件目录中列出所有上传的电子附件材料清单。

**（二）附件材料（逐项上传）**

上传的电子附件材料应为项目申请人和主要参与者取得的代表性成果或者科技奖励。

1．提供5篇以内申请人本人发表的与申请项目相关的代表性论文电子版文件；

2．如上传专著，可以只提供著作封面、摘要、目录、版权页等；

3．如上传所获科技奖励，应提供国家级科技奖励（国家自然科学奖、国家发明奖、国家科学技术进步奖）、省部级奖励（二等以上）奖励证书的电子版扫描文件；

4．如上传专利或其他公认突出的创造性成果或成绩，应提供证明材料的电子版扫描文件；

5．在国际学术会议上作大会报告、特邀报告，应提供邀请信或通知的电子版扫描文件；

6．根据项目申请的需要，附件材料**还可能**包含以下电子版扫描文件：在职攻读研究生学位的申请人的导师同意函、在站博士后申请人的依托单位承诺函、不具有高级专业技术职务（职称）且不具有博士学位申请人的推荐函、无工作单位或所在单位不是依托单位的申请人与申请项目依托单位签订的书面合同、依托单位非全职聘用的境内外人员的聘任合同复印件和相关说明材料、伦理委员会证明、加盖依托单位公章的国家社会科学基金结项证书复印件、依托单位生物安全保障承诺等。**具体要求参见本年度《国家自然科学基金项目指南》“申请须知”部分和正文“面上项目”部分相关科学部要求**。

**特别提示：上述附件第6项还需提供纸质原件，随纸质《申请书》一同报送。**