

# 谢海华

北京雁栖湖应用数学研究院 助理研究员

地址：北京市怀柔区河防口村 544 号 101408

电子邮箱：[haihua.xie@bimsa.cn](mailto:haihua.xie@bimsa.cn), [haihuaxie@foxmail.com](mailto:haihuaxie@foxmail.com) 移动电话：18710156236

## 研究兴趣

- 专业知识加工，知识挖掘及知识服务，知识图谱及知识库构建，知识库数据校验
- 文本大数据分析（包括媒体大数据分析、学术大数据分析、舆情分析、互联网信息检索等）
- 软件系统用户行为数据分析、用户意图推测和用户新需求挖掘

## 教育背景

- 2015.07~2017.05 北京大学 人工智能 博士后
- 2009.08~2015.05 美国爱荷华州立大学 计算机科学 博士
- 2006.09~2009.07 清华大学 软件工程 硕士
- 2002.09~2006.07 南京邮电大学 电子信息科学与技术 理学学士

## 工作经历

- 2021.10~至今 北京雁栖湖应用数学研究院 助理研究员
- 2017.06~2021.09 北大方正集团数字出版技术国家重点实验室 高级算法研究员、  
知识服务方向负责人、  
博士后导师
- 2015.07~2017.05 北大方正集团有限公司&北京大学 算法研究员、博士后
- 2009.08~2015.05 美国爱荷华州立大学计算机系软件工程实验室 研究助理

## 论文与论著

### ▪ 期刊论文

- [1] [Haihua Xie](#), Jingwei Yang, Carl K. Chang, Lin Liu, "A statistical analysis approach to predict user's changing requirements for software service evolution", *Journal of Systems and Software*, Volume 132, October 2017: 147-164. (SCI, JCR Q1)
- [2] Jianwei Li, Yongtao Wang, [Haihua Xie](#), Kai-Kuang Ma: Learning a Single Model With a Wide Range of Quality Factors for JPEG Image Artifacts Removal. *IEEE Trans. Image Process.* 29: 8842-8854 (2020). (SCI, JCR Q1)
- [3] [谢海华](#), 李奥林, 李亚博, 陈志优, 程静, 吕尚庆, 汤帆. CPLM-CSC: 基于单字级别预训练语言模型的中文错别字纠正方法[J]. *中文信息学报*, 2021, 35(5): 38-45.

- XIE Haihua**, LI Aolin, LI Yabo, CHEN Zhiyou, CHENG Jing, Lyu Xiaoqing, TANG Zhi. CPLM-CSC: Character-based Pre-trained Language Model for Chinese Spelling Checking and Correction. *Chinese Journal of information*, 2021, 35(5): 38-45.
- [4] **谢海华**,陈雪飞,都仪敏,吕肖庆,汤帆. 结合统计特征和图模型的半监督式中文关键短语抽取方法[J]. *中文信息学报*, 36 卷 4 期, 2022 年 4 月: 57-65.
- XIE Haihua**, CHEN Xuefei, DU Yimin, Lyu Xiaoqing, TANG Zhi. A Semi-supervised Method for Chinese Key Phrase Extraction Based on Statistical Features and Graph Model. *Chinese Journal of information*, 2022, 36(4): 57-65.
- [5] **谢海华**,陈志优,程静,吕肖庆,汤帆. 基于数据增强和多任务特征学习的中文语法错误检测方法[J]. *中文信息学报*, 36 卷 12 期, 2022 年 12 月: 16-23.
- XIE Haihua**, CHEN Zhiyou, CHENG Jing, LYU Xiaoqing, TANG Zhi. Chinese Grammar Error Detection based on Data Enhancement and Multi-task Feature Learning. *Chinese Journal of information*, 2022, 36(12): 16-23.

## ■ 会议论文

- [1] Lin Liu, **Haihua Xie**, Jingwei Yang, Yanni Li: Building Smart Systems to Satisfice User Intention in Situ: A Look Back to the Future. *IEEE International Conference on Software Services Engineering, SSE 2023 (SSE 2023)*: 262-271.
- [2] **Haihua Xie**, Xuefei Chen, Xiaoqing Lyu, Lin Liu: A Generalized Strategy of Chinese Grammatical Error Diagnosis Based on Task Decomposition and Transformation. *China Conference on Knowledge Graph and Semantic Computing, CCKS 2023*: 135-147.
- [3] **Haihua Xie**, Xiaoqing Lyu, Xuefei Chen. String Editing based Chinese Grammatical Error Diagnosis. *Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022)*, pages 5335-5344, October 12-17, 2022.
- [4] 张涛,汤珂,**谢海华**. 基于自然语言处理的高频经济学家信心指数构建和宏观预测. 第三届中国宏观经济学者论坛. 2022 年 6 月.
- [5] Jianwei Li, Yongtao Wang, **Haihua Xie**, Kai-Kuang Ma: Learning Local and Global Priors for JPEG Image Artifacts Removal. *IEEE Signal Process. Lett.* 27: 2134-2138 (2020).
- [6] Shaoling Jing, Shibo Hong, Dongyan Zhao, **Haihua Xie**, Zhi Tang: Combining Impression Feature Representation for Multi-turn Conversational Question Answering. *CNCL 2020*: 57-69.
- [7] Nuo Xu, **Haihua Xie**, Dongyan Zhao: A Novel Joint Framework for Multiple Chinese Events Extraction. *CNCL 2020*: 174-183.
- [8] **Haihua Xie**, Xuefei Chen, Yimin Du, Xiaoqing Lyu and Zhi Tang, "A Semi-supervised Method for Chinese Key Phrase Extraction based on Statistical Features and Graph Model", *Proceedings of The Fourteenth China Conference on Knowledge Graph and Semantic Computing (CCKS 2020)*.
- [9] **Haihua Xie**, Zhiyou Chen, Jing Cheng, Xiaoqing Lyu, Zhi Tang, "Chinese Grammar Error Detection based on Data Enhancement and Multi-task Feature Learning", *Proceedings of The Nineteenth China National Conference on Computational Linguistics (CCL 2020)*.

- [10] **Haihua Xie**, Zhiyou Chen, Yabo Li, Shaoling Jing, Xiaoqing Lyu, Zhi Tang, “SemBRS: A Semantic Analysis based Book Retrieval Approach”, *IEEE 3rd International Conference on Multimedia Information Processing and Retrieval (MIPR 2020)*.
- [11] **Haihua Xie**, Aolin Li, Yabo Li, Jing Cheng, Zhiyou Chen, Xiaoqing Lyu, Zhi Tang, “Automatic Chinese Spelling Correction based on Character-based Pre-trained Contextual Representations”, *Proceedings of The 8th CCF International Conference on Natural Language Processing and Chinese Computing (NLPC 2019)*.
- [12] **Haihua Xie**, Aolin Li, Yabo Li, Zhiyou Chen, Jing Cheng, Xiaoqing Lyu, Zhi Tang, “CPLM-CSC: Character-based Pre-trained Language Model for Chinese Spelling Checking and Correction”, *Proceedings of The Eighteenth China National Conference on Computational Linguistics (CCL 2019)*.
- [13] **Haihua Xie**, Xiaoqing Lyu, Xuefei Chen, Jinle Tong, Zhi Tang, “ES-ESens: Detection of Event Sentences based on Evaluation of the Explicitness and Significance of Information”, *Proceedings of IEEE 2nd International Conference on Multimedia Information Processing and Retrieval (MIPR 2019)*.
- [14] Yitao Liu, Xiaoqing Lyu, **Haihua Xie**, Xiaotong Yan, Bei Wang, Zhi Tang, “A Method for Improving the Reliability of Causal Inference from Large-scale Data in Biomedicine”, *Proceedings of 2018 International Conference on Bioinformatics and Biomedicine (BIBM18)*.
- [15] **Haihua Xie**, Xiaoqing Lyu, Zhi Tang, “Clustering Analysis-Based Approach to Detecting Entity Mixture in Knowledge Bases”, *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries (JCDL 2018)*: 395-396.
- [16] **Haihua Xie**, Xiaoqing Lyu, Zhi Tang, Xiaojun Huang, “Detection of Entity Mixture in Knowledge Bases Using Hierarchical Clustering”, *Proceedings of the 5th CCF Conference on Natural Language Processing and Chinese Computing and 24th International Conference on Computer Processing of Oriental Languages (NLPC-ICCPOL 2016)*: 288-299.
- [17] **Haihua Xie**, Xiaoqing Lyu, Zhi Tang, Mao Ye, “A Methodology to Evaluate Triple Confidence and Detect Incorrect Triples in Knowledge Bases”, *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries (JCDL 2016)*: 251-252.
- [18] **Haihua Xie**, Carl K. Chang, “Detection of New Intentions from Users Using the CRF Method for Software Service Evolution in Context-Aware Environments”, *Proceedings of the 39th Annual IEEE Computer Software and Applications Conference (COMPSAC 2015)*: 385-391.
- [19] **Haihua Xie**, Carl K. Chang, Hua Ming, Kai-Shin Lu, “The Concepts and Ontology of SiSL: A Situation-Centric Specification Language”, *Proceedings of the 36th Annual IEEE Computer Software and Applications Conference (COMPSAC 2012)*: 385-391.
- [20] Xiaoying Bai, Fevzi Belli, Elisa Bertino, Carl K. Chang, Atilla Elçi, Cristina Cerschi Secoleanu, **Haihua Xie**, Mohammad Zulkernine, “36th Annual IEEE Computer Software and Applications Conference”, *COMPSAC 2012*, Izmir, Turkey, July 16-20, 2012. IEEE Computer Society 2012, ISBN 978-1-4673-1990-4.

- [21] Xiaoying Bai, Fevzi Belli, Elisa Bertino, Carl K. Chang, Atilla Elçi, Cristina Cerschi Seceleanu, **Haihua Xie**, Mohammad Zulkernine, “36th Annual IEEE Computer Software and Applications Conference Workshops”, *COMPSAC 2012*, Izmir, Turkey, July 16-20, 2012. IEEE Computer Society 2012, ISBN 978-1-4673-2714-5.
- [22] **Haihua Xie**, “Observable Context-Based User Intention Specification in Context-Aware Environments”, *Proceedings of the 35th Annual IEEE Computer Software and Applications Conference (COMPSAC 2011)*: 712-715.
- [23] Tianying Li, Tong Li, **Haihua Xie**, Lin Liu, “i\*-Prefer: A Tool for Preference Model-Driven Decision Making”, *Proceedings of the 5th International i\* Workshop 2011 (iStar 2011)*: 151-153.
- [24] Wenting Ma, Lin Liu, **Haihua Xie**, Hongyu Zhang, Jinglei Yin, “Preference Model Driven Services Selection”, *Proceedings of the 21st International Conference on Advanced Information Systems Engineering (CAiSE 2009)*: 216-230.
- [25] **Haihua Xie**, Lin Liu, Jingwei Yang, “i\*-Prefer: Optimizing Requirements Elicitation Process Based on Actor Preferences”, *Proceedings of the 24th ACM Symposium on Applied Computing (ACM SAC 2009)*: 347-354.

## 授权专利&软件著作权

- [1] 发明名称: 信息价值评估方法、装置、设备及计算机可读存储介质  
 发明人: **谢海华**; 陈雪飞; 佟津乐  
 公开号: CN110866389B  
 授权日期: 2021.12.17  
 IPC分类: G06F40/289;G06F40/295;G06F40/253  
 CPC发明: G06F17/278;G06F17/2775;G06F17/274
- [2] 发明名称: 信息质量评估方法、装置、设备及计算机可读存储介质  
 发明人: **谢海华**; 陈雪飞; 佟津乐  
 公开号: CN110889274B  
 授权日期: 2022.02.08  
 IPC分类: G06F40/211;G06F40/295;G06F40/30  
 CPC发明: G06F17/271;G06F17/2785;G06F17/278
- [3] 发明名称: 信息明确度评估方法、装置、设备及计算机可读存储介质  
 发明人: **谢海华**; 陈雪飞; 佟津乐  
 公开号: CN110889289B  
 授权日期: 2022.05.06  
 IPC分类: G06F40/289;G06N3/04;G06F40/30  
 CPC发明: G06F17/2785;G06F17/2775;G06N3/04
- [4] 发明名称: 邮箱信息抽取方法及装置  
 发明人: **谢海华**; 罗学文; 陈雪飞; 佟津乐; 高良才; 黄肖俊; 汤帜  
 公开号: CN110020366B  
 授权日期: 2021.06.15

- IPC分类: G06F16/958;G06F16/903;G06F16/84  
 CPC发明: G06F17/30896;G06F17/2264
- [5] 发明名称: 抽取学者学术画像基本信息的方法与装置  
 发明人: 陈雪飞; 罗学文; [谢海华](#); 佟津乐; 黄肖俊; 高良才; 汤帜  
 公开号: CN110019761B  
 授权日期: 2021.04.30  
 IPC分类: G06F16/9535;G06F16/953;G06F16/335;G06F16/951  
 CPC发明: G06F17/30699;G06F17/30864
- [6] 发明名称: 检验知识库三元组的方法及装置  
 发明人: [谢海华](#); 黄肖俊; 吕肖庆; 汤帜  
 公开号: CN108021595B  
 授权日期: 2020.07.14  
 IPC分类: G06F16/30  
 CPC发明: G06F17/30731;G06F17/30705
- [7] 发明名称: 知识库同名实例混淆检测的方法及装置  
 发明人: [谢海华](#); 黄肖俊; 吕肖庆; 汤帜  
 公开号: CN108009184B  
 授权日期: 2021.08.27  
 IPC分类: G06F16/35;G06F16/36  
 CPC发明: G06F17/30734
- [8] 软件名称: The Cooperative Research Environment (CoRE) System  
 著作权人: Iowa State University  
 开发者: [Haihua Xie](#)  
 著作权类型: GNU General Public License
- [9] 软件名称: 主体偏好需求建模软件 (i\*-Prefer Modeller)  
 著作权人: 清华大学  
 开发者: 刘璘、[谢海华](#)  
 登记号: 2009SRBJ6677

## 技术应用

- 华服志—中华服饰文化云平台 (<http://www.huafuzhi.com>)
- 方正智能审校系统 (<http://book.founderss.cn>)
- 万方数据知识服务平台 (<https://www.wanfangdata.com.cn>)
- 清华大学出版社文泉学堂和文泉书局 (<https://www.bookask.com/>)

## 学术活动以及服务

- ICDH' 2024: IEEE International Conference on Digital Health Program Chair
- ICDH' 2023: IEEE International Conference on Digital Health Program Committee Member
- INDERSCIENCE Publishers Program Committee Member
- IEEE Services 2018 Congress Program Committee Member

- COMPSAC 2015 Conference MyReview Chair
- COMPSAC 2012 Conference Proceedings Chair
- COMPSAC 2011 Conference Proceedings Chair
- 中国计算机学会(CCF) 专业会员

## 指导的博士后课题

- 景少玲：基于阅读理解的多轮对话问答研究（2019.4~2021.3）
- 许诺：基于媒体数据的事件抽取与演化分析研究（2019.4~2021.3）
- 李健伟：基于深度学习的印刷物图像增强技术研究（2018.10~2020.9）

## 参加学术会议、论坛、报告及研讨会

1. 时间：2016年6月20日  
活动：16th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2016), 美国纽瓦克  
演讲：“A Methodology to Evaluate Triple Confidence and Detect Incorrect Triples in Knowledge Bases”
2. 时间：2016年12月6日  
活动：The Fifth Conference on Natural Language Processing and Chinese Computing (NLPC 2016), 云南昆明  
演讲：“Detection of Entity Mixture in Knowledge Bases Using Hierarchical Clustering”
3. 时间：2017年10月24日  
活动：中关村海淀园组织的第三期海英人才沙龙  
演讲：“Big data driven knowledge service applications”（特邀嘉宾报告）
4. 时间：2017年11月11日  
活动：The Sixth Conference on Natural Language Processing and Chinese Computing (NLPC 2017), 辽宁大连  
演讲：“Big data driven knowledge service applications”（特邀嘉宾报告）
5. 时间：2017年12月5日  
活动：中关村海淀园组织的博士后专场招聘会  
演讲：“人工智能驱动的知识服务发展”（博士后代表报告）
6. 时间：2018年7月7日  
活动：中国计算机学会组织的题为“AI 在智慧媒体领域的应用”的技术前线研讨会  
演讲：“知识服务中的人工智能应用”（特邀讲师报告）
7. 时间：2018年8月30日  
活动：智慧媒体研讨会, The Seventh CCF International Conference on Natural Language Processing and Chinese Computing (NLPC 2018), 内蒙古呼和浩特  
演讲：“Application and Innovation of Knowledge Services in Founder Digital Publishing Lab”（特邀嘉宾报告）
8. 时间：2018年12月10日  
活动：中关村海淀园组织的博士后专场招聘会  
演讲：“知识服务发展趋势”（博士后及博士后导师代表报告）

9. **时间:** 2019年3月21日  
**活动:** 2019年新疆图书馆数字资源建设研讨会, 新疆乌鲁木齐  
**演讲:** “人工智能技术驱动的数字资源服务创新”(特邀嘉宾报告)
10. **时间:** 2019年3月28号  
**活动:** IEEE 2nd International Conference on Multimedia Information Processing and Retrieval (IEEE MIPR 2019), 美国圣何塞  
**演讲:** “ES-ESens: Detection of Event Sentences based on Evaluation of the Explicitness and Significance of Information”
11. **时间:** 2019年8月21日  
**活动:** 第九届中国数字出版博览会  
**演讲:** “基于NLP技术的中文智能审校研究”(演讲嘉宾报告)
12. **时间:** 2019年10月12日  
**活动:** Workshop: Natural Language Understanding & Applications, The 8th CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC2019), 甘肃敦煌  
**演讲:** “NLP based Intelligent Text Proofreading for Chinese”(演讲嘉宾报告)
13. **时间:** 2019年11月26日  
**活动:** “遇见·预见”——2019年方正集团博士后高峰论坛暨方正集团博士后工作站建站20周年纪念活动  
**内容:** “NLP推动的数字出版技术革新”(演讲嘉宾报告)
14. **时间:** 2020年8月19日  
**活动:** 新一代数字技术对科技期刊出版业态的影响学术研讨会  
**内容:** “人工智能技术驱动数字出版和知识服务转型”(演讲嘉宾报告)
15. **时间:** 2020年10月18日  
**活动:** The 9th CCF International Conference on Natural Language Processing and Chinese Computing(NLPCC2020) - Industrial Speeches, 河南郑州  
**演讲:** “NLP Technology based Knowledge Service Applications”(演讲嘉宾报告)
16. **时间:** 2020年11月15日  
**活动:** 2020 China Conference on Knowledge Graph and Semantic Computing (CCKS 2020) – Poster/Demo, 江西南昌  
**演讲:** “结合统计特征和图模型的半监督式中文关键短语抽取方法”
17. **时间:** 2023年8月26日  
**活动:** 2023 China Conference on Knowledge Graph and Semantic Computing (CCKS 2023)  
**演讲:** “A Generalized Strategy of Chinese Grammatical Error Diagnosis based on Task Decomposition and Transformation”
18. **时间:** 2023年12月15日  
**活动:** 第二届学院路校友科技创新赛启动仪式暨科创企业家系列沙龙认知大模型与技术应用  
**演讲:** “中国 AI 大模型应用趋势解读”(主题演讲)

## 科研项目经历

- 时间：2019.12-2023.07  
项目：国家重点研发计划-专业内容知识聚合服务技术研发与创新服务示范（1200万元）  
编号：2019YFB1406300  
职责：负责子课题“跨媒体跨领域的资源标签自动生成与语义标注技术”
- 时间：2018.07-2021.07  
项目：基于出版资源内容的数据挖掘与知识提取技术研究（300万元）  
职责：主要负责人
- 时间：2015.07-2017.07  
项目：领域本体库置信度计算模型的研究和实现（20万元）  
职责：主要负责人
- 时间：2015.07-2017.07  
项目：中华数字书苑学知搜索系统（500万元）  
职责：学知搜索系统的主要负责人
- 时间：2016.01-2017.05  
项目：华服志-中华服饰文化云平台（500万元）  
职责：资源标签生成和资源关联度计算的负责人
- 时间：2009.01-2010.12  
项目：国家自然科学基金项目-基于社交建模的安全需求工程技术研究  
职责：参与
- 时间：2008.09-2010.07  
项目：973计划-面向特定应用的典型业务仿真技术与网构试验床  
职责：参与
- 时间：2007.09-2009.07  
项目：863计划-基于动态最优决策的需求工程技术研究  
职责：参与

## 技术获奖

- 2018年 CCL 2018技术评测-客服领域用户意图分类 二等奖（第二名）
- 2017年 2017开放学术精准画像大赛 第一阶段第二，第二阶段第五

## 奖励与荣誉

- 2018年 北京市第十三批海外高层次人才、北京市特聘专家（北京市千人计划-青年项目）
- 2019年 中关村高端领军人才（中关村高聚工程）
- 2019年 优秀博士后（一共20名） 中关村科技园区海淀区企业博士后科研工作站20周年
- 2001年 江西省“三好学生”