

耿光刚

耿光刚，博士，研究员，中国互联网络信息中心基础技术实验室副主任(主持工作)，中国反钓鱼网站联盟实验室主任，大数据流通与交易技术国家工程实验室数据安全与隐私保护研究中心主任，工信部网络安全试点示范项目负责人，中国科学院大学硕士生导师。研究方向包括：机器学习、信息检索、网络安全等。主持国家级项目 5 项(NSFC 项目 3 项，中央网信办专项 1 项，工信部项目 1 项)，参与发改委、科技部、基金委、中科院等项目 10 余项。发表学术论文 60 余篇(SCI/EI 索引 40 余篇)，起草国内外标准近 20 项(其中 6 项已发布或报批)，申请专利 50 余项(获授权 11 项，含 1 项 PCT 专利获美国专利局授权)。先后 4 次带队参加互联网信息检测、评估、检索等国际学术竞赛(包括: Web Spam Challenge 2007、Web Spam Challenge 2008、ECML/PKDD 2010 Discovery Challenge 和 CCF NLP Challenge 2016 on Entity Search)，均获得冠军。担任 QShine2016、ICCE2017、VTC2017-Spring 等程序委员会委员,担任CAA模式识别与机器智能专业委员会委员、CAAI模式识别专业委员会委员。

研究方向

机器学习、Web 信息检索、互联网大数据分析、网络安全

个人经历

2016-至今	中国互联网络信息中心	研究员，基础技术实验室副主任（主持工作）
2010-2016	中国科学院计算机网络信息中心	副研究员，国家域名安全中心主任助理
2008-2010	中国科学院计算机网络信息中心	助理研究员
2005-2008	中国科学院自动化研究所	模式识别与智能系统专业，工学博士

课题承担情况

2016.12 至今	电信和互联网行业网络安全试点示范项目（工信部）	项目负责人
课题名称：网络钓鱼发现与处置一体化服务平台		
2014.01 至今	国家自然科学基金面上项目（NSFC）	项目负责人
课题名称：复杂环境下 Web 内容质量评价的统计学习方法研究		
2014.11-2015.12	网络安全与信息化专项（中央网信办）	执行负责人
课题名称：打击“死而复生”违法违规网站专项		
2011.01-2013.12	国家自然科学基金青年基金项目（NSFC）	项目负责人
课题名称：多模态 Web 作弊检测的统计机器学习方法研究		
2013.01-2013.12	国家自然科学基金国际交流与合作项目（NSFC）	项目负责人
课题名称：中国-东南亚互联网基础资源战略合作峰会		
2014.01-2015.12	中科院网络中心 135 重点培育方向专项（中科院）	子课题负责人
课题名称：可信的互联网基础资源服务体系及安全保障技术研究		
2007.07-2011.12	973 项目（科技部）	主要参与人
课题名称：基于非结构化信息处理的综合决策与评估		

竞赛获奖情况

CCF Challenge 2016 on NLP and Intelligent Analysis, Baidu Cup 冠军 团队联合负责人

- ✓ 背景：CCF NLP 国际挑战赛由百度与中国计算机学会（CCF）联合发起，面向全球自然语言处理领域的研究人员，致力于探讨通过人工智能解决实际互联网大数据分析问题。本次挑战赛主题为“Search proper entities for a given query”，分四个不同的数据分析任务，属于典型的互联网大数据分析，174 支参赛团队来自清华大学、北京大学、中国科学院、伦敦大学等 70 多个学校和科研机构。
- ✓ 概况：作为参赛团队联合负责人，负责“名人”和“餐厅”搜索任务（Dr X.B. Jin 负责“电影”和“电视”

任务)。设计数据采集、特征提取、模型设计等研究方案，并具体实施。所提交的结果获得总成绩第一名，并获得 **25,000 元** 现金奖励和 Travel Grant，本次挑战赛由 **Baidu** 提供奖金。

✓ 网址：<http://nlpcc.baidu.com/finalresult.html>

ECML/PKDD 2010 Discovery Challenge 冠军 团队负责人

✓ 背景: Discovery Challenge 是机器学习和数据挖掘领域最富盛名测评之一,2010 年的测评是连续的第 12 届。ECML/PKDD2010 任务挑战赛的主题为 Web 内容质量评价(查询无关排序)和网页内容自动分类。数据集包含约 2380 万网页, 6 亿超链接。由 **Yahoo**、**Google**、欧洲档案馆和匈牙利科学院等联合主办。

✓ 概况: 作为参赛团队负责人, 设计研究方案, 并负责具体实施, 包括特征提取、排序算法等, 提交的结果在英语、跨语言(法语、德语)、分类任务以及总成绩均获得第一名, 其中总成绩超出第二名 **1.4%**, 并获得 **2,500 美元** 现金奖励, 本次挑战赛由 **Google** 和 **Yahoo** 提供奖金。

✓ 网址：<http://www.ecmlpkdd2010.org/indexd7fa.html?md=articles&id=2041&lg=eng>

Web Spam Challenge 2007 和 Web Spam Challenge 2008 冠军 团队负责人

✓ 背景: 搜索引擎作弊是互联网搜索目前面临的巨大挑战之一, 体现在作弊的多样性、隐蔽性、融合性、进化性等方面。Web Spam Challenge 系列竞赛由 Yahoo 欧洲研究院发起, 吸引了全球范围内的对抗性信息检索研究人员参与, 2007 年和 2008 年两次测评的参赛队来自世界各地, 包含 UC 伯克利大学、微软研究院、雅虎研究院、法国电信等。

✓ 概况: 作为负责人于 2007 年和 2008 年两次参与该项国际竞赛, 均取得第一名的成绩, 其中 2008 年的测评结果 (AUC) 超过第二名 **2.4%**。

✓ 网址：<http://webspam.lip6.fr/wiki/pmwiki.php>

论文情况

部分期刊论文

1. X.B. Jin, **G.G. Geng**, K.Z. Huang, and Z.W. Yan. SEER: Semantic Entity Ranking with Comprehensive Knowledge. *IEEE Transactions on Systems, Man and Cybernetics: Systems*. (Submitted)
2. X.B. Jin, **G.G. Geng**, G.S. Xie, K.Z. Huang. Pair-wise Loss for Optimizing NDCG Approximately. *Information Sciences*. (Submitted)
3. Y.L. Liao, W.W. Chen, K.C. Wu, D.Y. Li, X. Liu, **G.G. Geng**, Z. Su and Z. Zheng. A site selection method of DNS using the particle swarm optimization algorithm. *Transactions in GIS*. 2016. doi:10.1111/tgis.12244 (SCI) (Accepted)
4. J.L. Zhang, Z.W. Yan, **G. G. Geng***, X.D. Lee, and J.H. Lee. Trademark Protection for Chinese Domain Names. *Journal of Internet Technology*. (通讯作者) (SCI) (Accepted)
5. Z.W. Yan, S. Zeadally, **G. G. Geng***. ISBORD: Internet Searching based on Resource Description. *ICT Express*. (通讯作者) (Accepted)
6. Z.W. Yan, **G. G. Geng***, S. Zeadally, Y.J. Park. Distributed All-IP Mobility Management Architecture Supported by the NDN Overlay. *IEEE Access*. (通讯作者) (SCI) (Accepted)
7. J. Jin, Z.W. Yan, J.H. Lee and **G.G. Geng**, Health Evaluation of a Domain Name System Based on the Analytic Hierarchy Process, *Journal of Internet Technology*. January 2018, Vol.19 No.1. (SCI)
8. S.M. Wang, Z.W. Yan, **G.G. Geng**, and Y.D. Zhang. Geo-based Content Naming and Forwarding Mechanism for Vehicular Networking over CCN. *Int. J. of Internet Technology and Secured Transactions*. Vol. 6, No. 4, 2016, Pages 291-302.
9. **G.G. Geng***, X.D. Lee. Combating Phishing Attacks via Brand Identity & Authorization Features. *Security and Communication Networks*. 2015. (通讯作者) (SCI)

10. X.B. Jin, **G.G. Geng***, Sun M, et al. Combination of Multiple Bipartite Ranking for Multipartite Web Content Quality Evaluation[J]. *Neurocomputing*, Vol 149, Part C, 3 February 2015, Pages 1305–1314 (通讯作者) (SCI)
11. Y.M. Zhang, K.Q. Huang, **G.G. Geng**, C.L. Liu. MTC: A Fast and Robust Graph-Based Transductive Learning Method. *IEEE Transactions on Neural Networks and Learning Systems*. 2014. (SCI)
12. X.C. Zhang, W.D. Gu, M.H. Yang, **G.G. Geng**, W.M. Luo. A distance-heuristic tree building approach in application layer multicast, *Computing and Informatics*, Vol.31, pp 1001-1030. 2013. (SCI)
13. **G.G. Geng***, L.M. Wang, W. Wang, S. Shen. Statistical Cross-Language Web Content Quality Assessment. *Knowledge-Based Systems*. Vol. 35, Nov 2012, 312-319. (通讯作者) (SCI)
14. X.C. Zhang, M.H. Yang, **G.G. Geng**, et al. A Two-Tiered Reliable Application Layer Multicast. *Computer Science and Information Systems*, Vol. 8, June 2011. (SCI)
15. X.C. Zhang, M.H. Yang, **G.G. Geng**, W.M. Luo. A DFISM-based protocol conformance testing and diagnosing method. *Informatica*, 22(3), 447-469. 2011. (SCI)

部分会议论文

16. Z.W. Yan, **G.G. Geng**, H. Nkazato, Y.J. Park, K. Nisar and A.A.A. Ibrahim. On-demand DTN Communications in Heterogeneous Access Networks based on NDN, *RAFNET2017*.
17. X.B. Yuchi, **G.G. Geng**, Z.W. Yan, and X.D. Lee. Towards Tackling Privacy Disclosure Issues in the DNS. *IFIP/IEEE Integrated Network Management Symposium (IM 2017)*.
18. X.N. Li, **G.G. Geng**, Z.W. Yan, Y. Chen and X.D. Lee. Phishing Detection Based on Newly Registered Domains. *3rd International Workshop on Privacy and Security of Big Data*, December 8 2016.
19. X.B. Jin, **G.G. Geng**, K.Z. Huang and Z.W. Yan. Statistical Entity Ranking with Domain Knowledge. *24th International Conference on Computer Processing of Oriental Languages*. 2016: 811-818.
20. Z.W. Yan, **G.G. Geng**, X.D. Lee and Y.J. Park, A Novel DMM Architecture based on NDN, *EAI QSHINE*, Seoul, Korea, July 7-8, 2016.
21. C.C. Wang, **G.G. Geng**, and Z.W. Yan, A Cross-Domain Hidden Spam Detection Method based on Domain Name Resolution, *EAI QSHINE*, Seoul, Korea, July 7-8, 2016.
22. X.W. Liu, Z.W. Yan, **G.G. Geng**, et al. RPKI Deployment: Risks and Alternative Solutions. *Genetic and Evolutionary Computing*. Springer International Publishing, 2016: 299-310.
23. **G.G. Geng**, X.T. Yang, W. Wang, C.J. Meng. A Taxonomy of Hyperlink Hiding Techniques. *APWeb 2014*. 165-176.
24. **G.G. Geng**, X.D. Li, W. Wang, S.S. Tseng. Favicon -- a Clue to Phishing Sites Detection. 8th IEEE *eCrime Researchers Summit 2013*.
25. Y.M. Zhang, K.Z. Huang, **G.G. Geng**, et al. Fast Knn Graph Construction with Locality Sensitive Hashing. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 2013: 660-674.
26. X.B. Jin, **G.G. Geng**, J.W. Yu, D.X. Zhang. Multi-label Learning Vector Quantization Algorithm, *ICPR'2012*.
27. X.B. Jin, Yu, J., Zhang, D., & **G.G. Geng**. Multi-partite ranking with multi-class AdaBoost algorithm. In *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2012, 9th International Conference on (pp. 884-887).
28. **G.G. Geng**, X.D. Li, L.M. Wang, W. Wang, S. Shen. Statistical Feature Extraction for Cross-Language Web Content Quality Assessment, *34th Annual ACM SIGIR Conference, SIGIR'2011*.

29. **G.G. Geng**, Q.D. Li, X.C. Zhang. Link Based Small Sample Learning for Web Spam Detection, *18th International World Wide Web Conference, WWW'2009*.
30. **G.G. Geng**, C.H. Wang, Q.D. Li. Improving Web Spam Detection with Re-Extracted Features, *17th International World Wide Web Conference, WWW'2008*.
31. Q.D. Li, C.H. Wang, and **G.G. Geng**. Improving Personalized Services in Mobile Commerce by a Novel Multicriteria Rating Approach. *17th international conference on World Wide Web*. ACM, 2008.
32. **G.G. Geng**, C.H. Wang, Q.D. Li. Improving Spamdexing Detection via Two-Stage Classification Strategy, *Asia Information Retrieval Symposium, AIRS'2008*.
33. **G.G. Geng**, C.H. Wang, Q.D. Li, et al Fighting Link Spam with a Two-Stage Ranking Strategy, *29th European Conference on Information Retrieval, ECIR'2007*.
34. **G.G. Geng**, C.H. Wang, Q.D. Li, et al. Boosting the Performance of Web Spam Detection with Ensemble Under-sampling Classification. *Fuzzy Systems and Knowledge Discovery, 2007. Fourth International Conference on. IEEE*, 2007, 4: 583-587.

标准化情况

中国通信行业标准

1. **耿光刚**,延志伟,傅瑜,李洪涛,马军锋. 互联网码号资源公钥基础设施 (RPKI) RPKI与BGP路由器交互协议, YD/T 1229-2016, 报批稿
2. 延志伟,**耿光刚**,傅瑜,李洪涛,马军锋.互联网码号资源公钥基础设施 (RPKI) 体系结构,YD/T 1091-2016, 报批稿
3. 李晓东,胡安磊,**耿光刚**,陈勇,延志伟.钓鱼攻击举报数据交换协议技术要求,YD/T 3038-2015, 已发布
4. 傅瑜,延志伟,**耿光刚**,李洪涛,马军锋. 互联网码号资源公钥基础设施 (RPKI) 资料库架构,YD/T 1228-2016, 报批稿
5. 毛伟,李晓东,李洪涛,王奠,徐颖,**耿光刚**.域名注册系统安全防护要求,YD/T 2245-2011, 已发布
6. 毛伟,李晓东,李洪涛,王奠,徐颖,**耿光刚**.域名注册系统安全防护检测要求,YD/T 2246-2011, 已发布

IETF 国际标准

7. Z.W. Yan, X.W. Liu, X.D. Lee, **G.G. Geng**, Y. Fu. RPKI Deployment Considerations: Problem Analysis and Alternative Solutions, 个人草案

专利情况

获授权专利

1. **G.G. Geng**, and Bo Hong. Method for detecting phishing website without depending on samples. U.S. Patent No. 9,276,956. 1 Mar. 2016 (PCT 专利, 获美国专利商标局授权)

2. **耿光刚**,王利明,胡安磊. 基于域名解析特征的网站作弊检测方法, 授权公告号: CN103684896B, 授权日: 2017-02-01
3. **耿光刚**,洪博.一种不依赖样本的钓鱼网站检测方法, 授权公告号: CN103685174B, 授权日: 2016-12-21
4. **耿光刚**,王伟. 基于网页图标匹配的品牌仿冒网站检测方法, 授权公告号: CN103281320B, 授权日: 2016-12-07
5. **耿光刚**,王春恒等. 一种基于小样本集的搜索引擎作弊检测方法, 授权公告号: CN101350011B, 授权日: 2011-09-07
6. **耿光刚**,李秋丹,王春恒等. 一种搜索引擎作弊检测优化方法, 授权公告号: CN101493819B, 授权日: 2011-09-14
7. 王利明,**耿光刚**,洪博.不良网页检测方法及装置, 授权公告号: CN102523130B, 授权日: 2015-02-04
8. 洪博,**耿光刚**,王利明,肖雅丽. 网络钓鱼检测方法及装置, 授权公告号: CN102316099B, 授权日: 2014-10-22
9. 洪博,**耿光刚**,王利明.不良网站检测方法及设备, 授权公告号: CN102663093B, 授权日: 2014-07-09
10. 王利明,**耿光刚**,陈威. 网站内容质量评估方法和装置, 授权公告号: CN102243661B, 授权日: 2014-04-23
11. 肖雅丽,王利明,**耿光刚**,洪博. 钓鱼网站检测方法及装置, 授权公告号: CN102523210B, 授权日: 2014-11-05

注: 另有 30 项国内专利和 10 余项 PCT 专利处于审理阶段。

服务孵化

- **Web滥用检测**: 研发了具有自主知识产权的互联网黄赌毒枪爆等滥用信息的主动发现和判定系统, 日处理20亿级网页数据, 仅2016全年就主动发现并举报近6万个不良域名, 成为打击国家域名滥用的最重要途径, 极大节省了人力核查成本。
- **反网络欺诈**: 研发了钓鱼欺诈网站主动探测与识别平台, 已累计发现并处置钓鱼欺诈网站近10万个, 并为平安集团、京东商城等提供反钓鱼服务, 该服务成为我国打击网络钓鱼欺诈的最重要途径之一。该平台成功入选2016年电信和互联网行业网络安全试点示范项目。
- **基础资源库**: 通过域名注册解析数据、Web应用数据、网路拓扑探测数据等多维度分析, 构建并持续更新全球域名库、IP地理位置库和全球中文网站库等基础资源库, 为多种服务提供基础支撑。

社会服务

学会参与: IEEE会员, 中国自动化学会会员, 中国人工智能学会会员, 中国计算机学会高级会员

专委会服务: 中国自动化学会模式识别与机器智能专业委员会委员, 中国人工智能学会模式识别专业委员会委员, 中关村区块链产业联盟专家委员会委员, 大数据流通与交易技术国家工程实验室专家委员会委员

研究生指导: 中国科学院大学硕士生导师

论文评审: IEEE Transactions on Neural Networks and Learning Systems, Security and Communication Networks, Knowledge-Based Systems, Knowledge and Information Systems, Neurocomputing, Computers & Electrical Engineering

奖励&项目评审: 中宣部重大项目评估专家, 国家自然科学基金、北京市自然科学基金、山东省自然科学基金评审专家, 北京市科普专家, 北京市奖励评审专家

TPC Member: 12th EAI International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine2016), IEEE International Conference on Consumer Electronics (ICCE2017), IEEE 85th Vehicular

Technology Conference (VTC2017-Spring)

其他服务:中国反钓鱼网站联盟实验室主任, 大数据流通与交易技术国家工程实验室数据安全与隐私保护研究中心主任, 《网络与信息安全学报》特邀编辑