

DOMINIK WURZER

Ph.D in Computer Science, Austrian

✉ dominik@wurzer.com ☎ +86187-7110-7346 ✉ Flat 1801, Unit 6, East Lake Mansion, Wuhan
📍 Wuhan, China in linkedin.com/in/dominik-w-80032a4a/ 🗣 Domwu123



EXPERIENCE

Teaching Assistant

University of Edinburgh

📅 9.2013 – 12.2016

📍 Edinburgh, UK

I designed and held Tutorials, Lab-sessions, Coursework and Exams for Information Retrieval and Applied Machine Learning course for Master students.

My teaching included:

- Search Engine Development
- Social Media data stream processing and mining
- Big Data Analysis on Large Textual Data Set
- Automated Question Answering using the Web

Lecturer

Godfrey Okoye University

📅 7.2010 – 9.2010

📍 Enugu, Nigeria

I designed and held lectures for Master Students in Computer Science.

My courses included:

- Programming with C
- Programming with Python

Software Developer

hs2n Information Technology GmbH

📅 1.2010 – 7.2010

📍 Klagenfurt, Austria

- Implementation of network based administration tools for commercial applications
- Design and implementation of a fault-tolerant transfer platform that moves massive files over the internet

Soldier

Austrian Armed Forces

📅 6.2009 – 12.2009

📍 Austria

- Helicopter based Alpine Border Security
- Managing the Logistics and Transportation of Military Grade Materials

Research Assistant

Carinthia University of Applied Sciences

📅 6.2008 – 6.2009

📍 Villach, Austria

- Development of a robotic system (hard and software) for educational purpose

STRENGTHS

Resilient Focus Attention to details

Hard-working Ambitious

SQL Python HTML & CSS

Javascript PHP Java C#

Arduino C C++ JQuery .NET

JSON Keras TensorFlow

Hadoop-MapReduce

Windows Linux

LaTeX Office

EDUCATION

Ph.D in Computer Science

University of Edinburgh

📅 11.2012 – 9.2016

Thesis: Scaling First Story Detection to Massive Streams

Supervisor: Professor Victor Lavrenko & Professor Henry Thompson

External Adviser: Professor Miles Osborne (Bloomberg)

M.S.c in Artificial Intelligence and Cognitive Science

University of Edinburgh

📅 9.2011 – 8.2012

Focus: information retrieval, machine learning, large scale computing

Thesis-Topic: The Randomized Relevance Retrieval Model

Success: with distinction (honours, top 3%)

B.S.c in Computer Science

University of Central Lancashire

📅 9.2010 – 6.2011

Thesis-Topic: High-Performance data transfer web applications

Success: with distinction (honours, best student in 2010/2011)

RESEARCH PROJECTS

Event Detection in Social Media

EPSRC and DSTL, grant number EP/L010690/1 & EPSRC, grant number EP/J020664/1

📅 12.2012 - 12.2016

📍 Edinburgh, UK

The main goal of this project was to find all breaking news on Social Media, as quickly as possible. With 5.000 messages per second on Twitter, the challenges include developing machine learning methods that scale well and can detect breaking news in a sea of garbage.

The project is run by a collaboration between Prof. Miles Osborne and Prof. Victor Lavrenko from the School of Informatics at the University of Edinburgh, as well as Prof. Iadh Ounis, and Prof. Craig MacDonald from the School of Computer Science at the University of Glasgow.

My contributions:

Low Latency Web Crawling and Search Engine

- Major search engine providers only allow searching websites with a certain delay. For research purposes at the University of Edinburgh, I have designed and built low latency web crawlers and search engines. These allow tracking frequently and rapidly updated website content with a much smaller delay.

Social Media Crawler

- I have built robust social media crawlers for the University of Edinburgh for research data set. In particular, I built crawlers for Twitter and Sina Weibo, whose data was used in numerous publication by the University of Edinburgh.

Memory based algorithms & k-term hashing

- I invented memory based novelty detection and k-term hashing, which were used in the currently fastest published First Story Detection system.

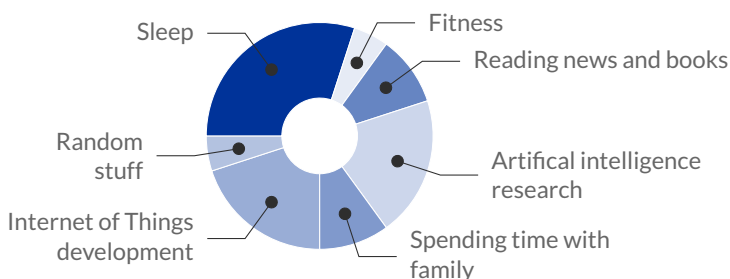
Graph Analysis

- I built analysis tools based on the HIT and PageRank algorithms to identify and visualize a corporation's key persons and their relations with each other's based on internal email communication.

Port decades old algorithms and evaluation paradigms to modern data sets

- I managed to apply decades old First Story Detection algorithms like the UMass systems to modern fast moving data streams. Additionally, I was able to port the original TDT evaluation scripts to run on Social Media data sets like the Edinburgh Twitter corpus and the Cross-Twitter Data Set.

A DAY OF MY LIFE



LANGUAGES

German



English



LIFE PHILOSOPHY

"Per ardua ad astra."

"Through hard work to the stars."

ACCOMPLISHMENTS

Memory based algorithms & k-term hashing

I invented k-term hashing, a high-performance novelty detection algorithm applicable to large scale streaming applications, which operates at a space & time complexity of $O(1)$

The worlds fastest First Story Detection algorithm

I developed the worlds fastest First Story Detection system, which exceeds state-of-the-art methods by 2 orders of magnitude without sacrificing accuracy

The first instantaneous Rumour Detection system

I built the first real-time rumour detection system for high volume social media streams including Twitter and Sina Weibo

Streaming Topic Tracking

I developed the first algorithm able to track a stream of topics against another fast-moving stream of documents

Randomized Relevance Model

I randomized the Relevance Model for information retrieval through Locality Sensitive Hashing (LSH), which reduces the runtime by 1 order of magnitude without sacrificing accuracy

PUBLICATIONS

Journal Articles

- Yumeng Qin, Dominik Wurzer, Victor Lavrenko, Cunchen Tang, "Predicting Future Rumour", Accepted by Chinese Journal of Electronics, SCI, 2018.

Conference Proceedings

- Yumeng Qin, Dominik Wurzer, Victor Lavrenko, Cunchen Tang, "Counteracting novelty decay in first story detection", *European Conference on Information Retrieval*, Lecture Notes in Computer Science, Springer, Aberdeen, UK, Vol. 10193, pp.555–560, 2017.
- Dominik Wurzer, Victor Lavrenko and Miles Osborne, "Tracking unbounded Topic Streams", *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, ACL, Beijing, China, Vol. 1, pp.1765–1773, 2015.
- Dominik Wurzer, Victor Lavrenko and Miles Osborne, "Twitter-scale New Event Detection via K-term Hashing", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, pp.2584–2589, 2015.

MOST PROUD OF

Austrian National Champion for Hexapod Racing

I won the Austrian national championship for constructing and programming and racing a hexapod robot, while I attended High School.

Youngest Lecturer at a University in Nigeria

At age 20, I was the youngest lecturer in Nigerian to teach a course in Computer Science at a university.

Overstepped the first two years of my BSc. for Computer Science

Following outstanding academic achievements during my time at High School in Austria, I overstepped the first two years of my Bachelor of Science at the University of Central Lancashire, UK.

Best student for Computer Science award

I was recognized as the best student in Computer Science at the University of Central Lancashire for the year of 2010/2011.

Scholarship of Excellence

I won the Scholarship of Excellence, awarded by the Austrian government for academic achievements during my Bachelor in Computer Science.

PhD Scholarship

Following repeated outstanding academic achievements during my BSc. at the University of Central Lancashire and my MSc. at the University of Edinburgh, I was awarded with a SICSA funding to attend my PhD program at the University of Edinburgh. I was 1 of only 5 persons to be awarded this prize in 2012.

RESEARCH INTERESTS

Real-time streaming algorithms

Data mining on social media streams

Applied machine learning

Information retrieval

Natural language processing

Randomized algorithms

DOMINIK WURZER

博士研究生 -- 计算机科学, 奥地利人

@ dominik@wurzer.com (+86) 187-7110-7346 湖北省武汉市洪山区八一路340号东湖楚天府6栋1801
📍 中国, 武汉 in linkedin.com/in/dominik-w-80032a4a/ 📧 Domwu123



工作经历

教学助理

爱丁堡大学

📅 2013年9月 -- 2016年12月 📍 英国, 爱丁堡

针对硕士生开设的信息检索与应用机器学习课程进行教学辅助, 主要职责包括上机课程、辅导课程、作业及考试的设计、辅导与教学。

教学内容:

- 1、搜索引擎开发
- 2、社交媒体数据流处理与挖掘
- 3、大文本数据集的大数据分析
- 4、利用Web自动答疑

讲师

Godfrey Okoye大学

📅 2010年7月 -- 2010年9月 📍 尼日尼亚, 埃努古

针对计算机科学系的硕士生进行课程设计与教学, 主要课程包括:

- 1、C语言编程
- 2、Python语言编程

软件开发

hs2n Information Technology GmbH

📅 2010年1月 -- 2010年7月 📍 奥地利, 克拉根福

- 1、基于网络的商业应用管理工具的开发
- 2、将大量文件进行互联网传输的容错传输平台的设计与开发

军人

奥地利军队

📅 2009年6月 -- 2009年12月 📍 奥地利

- 1、阿尔卑斯山脉边境安全 (直升机巡逻)
- 2、管理军用级物料的物流和运输

助理研究员

克恩顿州应用科技大学

📅 2008年6月 -- 2009年6月 📍 奥地利

基于教学目的的机器人系统的开发 (软硬件)

优势及专业技能

弹性/恢复性强 | 注意力集中 | 注重细节
勤奋努力 | 壮志凌云

SQL | Python | HTML & CSS | Javascript | PHP
Java | C# | Arduino | C | C++ | JQuery | .NET
JSON | Keras | TensorFlow
Hadoop-MapReduce

Windows | Linux

LaTeX | Office

教育经历

博士研究生 -- 计算机科学

爱丁堡大学

📅 2012年11月 -- 2016年9月

毕业论文: 大规模数据流中的新话题检测

指导教师: Victor Lavrenko教授、Henry Thompson教授

外部指导: Miles Osborne教授

硕士研究生 -- 人工智能与认知科学

爱丁堡大学

📅 2011年9月 -- 2012年8月

主修: 信息检索、机器学习、大规模计算

毕业论文题目: 随机相关检索模型

成就: 杰出 (荣誉学位)

本科学士 -- 计算机科学

中央兰开夏大学

📅 2010年9月 -- 2011年6月

毕业论文题目: 高性能数据传输web应用

成就:

- 1、杰出 (荣誉学位)
- 2、2010/2011年度最佳学生奖

项目经验

社交媒体事件检测

EPSRC and DSTL, grant number EP/L010690/1 & EPSRC, grant number EP/J020664/1

2012年12月 -- 2016年12月

英国, 爱丁堡

该项目的主要目标是尽快地在社交媒体数据流中检测所有的突发新闻。然而, 社交媒体巨头 -- Twitter平台中每秒约有5000条消息产生, 使得本项目挑战性在于保证在垃圾数据海洋中检测到突发事件的准确率, 同时还须保证该机器学习方法的可扩展性。

该项目属于合作型科研类项目, 由爱丁堡大学信息学院的Miles Osborne及Victor Lavrenko教授与格拉斯哥大学计算机学院的Ladh O-unis及Craig MacDonald教授共同开展。

本人贡献:

低延迟网页抓取与搜索引擎

现有主流搜索引擎供应商提供的网站搜索都存在一定程度的延迟。基于在爱丁堡大学开展科研的目的, 本人设计和构建了低延迟的网络爬虫和搜索引擎, 能够在更小的延迟内跟踪频繁和快速更新的网站内容。

社交媒体爬虫工具

本人为爱丁堡大学的研究数据收集开发了强大的社会媒体的爬虫工具, 包括基于Twitter和新浪微博的两款爬虫工具。通过这两款工具所采集的数据集在爱丁堡大学的多篇发表文献中都有应用。

基于记忆的算法及Kterm Hashing算法

本人提出了基于记忆的新颖度检测概念, 并在此基础上设计实现了Kterm hashing算法, 该算法被目前所有公开的话题检测系统中最快速的检测系统所采用。

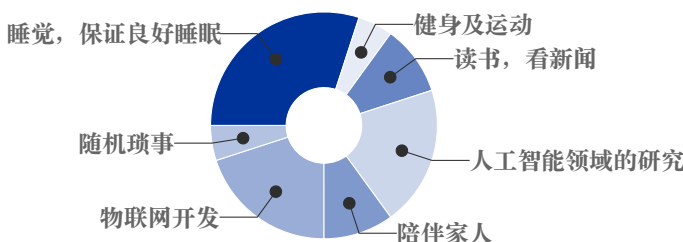
图分析

本人开发了基于HIT和PageRank算法的图分析工具, 该工具能够在公司内部邮件系统的基础上识别该公司的核心人物及其相互关系, 并将分析结果可视化。

将陈旧的算法及评估范式拓展到现代数据集中

本人将本世纪初提出的新话题检测算法 (如UMass) 拓展到基于现代快速移动数据流的检测任务中来。此外, 本人对原TDT评估脚本进行拓展, 使其能够应用在现代社交媒体数据集中, 如爱丁堡大学的Twitter语料库及Cross-Twitter数据集。

工作日例程



语言技能

德语



英语



人生观

“Per ardua ad astron.”

“宝剑锋自磨砺出, 梅花香自苦寒来。”

主要成果

基于记忆的算法 & k-term hashing算法

发明了Kterm hashing算法, 该算法能够在大规模数据流中进行新颖度检测, 且其时间与空间复杂度为O(1)。

最快的新话题检测算法

研发了最快速的新话题检测算法, 该算法能够在保证准确率的同时拥有超越其他检测算法两个量级的文档处理速度。

谣言实时检测系统

构建了第一个能够在海量社交媒体数据流中 (包含Twitter和新浪微博) 实时地进行谣言检测的系统。

流话题跟踪

开发了第一个能够跟踪在快速移动的文件流中对一连串的话题流进行跟踪的算法。

随机相关模型

利用局部敏感哈希 (LSH) 将信息检索任务中的相关模型随机化, 成功地在保证准确率的同时将运行时间降低了一个量级。

论文发表

📖 期刊文章

1. Yumeng Qin, Dominik Wurzer, Victor Lavrenko, Cunchen Tang, “Predicting Future Rumour”, Accepted by Chinese Journal of Electronics, SCI, 2018.

👥 会议文章

1. Yumeng Qin, Dominik Wurzer, Victor Lavrenko, Cunchen Tang, “Counteracting novelty decay in first story detection”, European Conference on Information Retrieval, Lecture Notes in Computer Science, Springer, Aberdeen, UK, Vol. 10193, pp.555--560, 2017.
2. Dominik Wurzer, Victor Lavrenko and Miles Osborne, “Tracking unbounded Topic Streams”, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, ACL, Beijing, China, Vol. 1, pp.1765--1773, 2015.
3. Dominik Wurzer, Victor Lavrenko and Miles Osborne, “Twitter-scale New Event Detection via K-term Hashing”, Proceedings of the Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, pp.2584--2589, 2015.

我的骄傲

🏆 六足机器人竞赛奥地利全国冠军赛
读中学时，构建和编程的六足机器人在奥地利全国锦标赛中获得冠军。

👤 尼日利亚某大学的最年轻讲师
20岁时，成为了尼日利亚某大学最年轻的讲师，完成了计算机科学课程的教学与辅导。

🏛️ 攻读计算机专业理学学士学位期间，被批准直接进入第三年学习
凭借在奥地利攻读高中期间的卓越成绩，被英国中央兰开夏大学批准免修本科前两年学分，直接进入第三年学习。

😊 计算机科学专业的最佳学生奖 (2010/2011)
被授予英国中央兰开夏大学计算机专业2010-2011年度最佳学生奖荣誉。

€ 卓越奖学金 (2011)
在攻读学士学位期间，获得了由奥地利政府颁发的卓越奖学金。

🎓 博士奖学金
凭借本科和硕士期间的优异成绩和学术成就，被授予了攻读爱丁堡大学博士学位的奖学金(SICSA)。该奖学金由苏格兰政府颁发且每年度仅选取5个优秀学生进行资助。

研究方向 & 兴趣

实时流算法

社交媒体数据挖掘

应用机器学习

信息检索

自然语言处理

随机算法