José Devezas, PhD

data \rightarrow information \rightarrow knowledge \rightarrow search & analytics



Contact Porto, Portugal

joseluisdevezas@gmail.com https://josedevezas.com LinkedIn: jldevezas

Languages

Portuguese (native) English (fluent)

Programming

Python, R, SQL, Scala, Java, Gremlin, Vue.js

Education

2016-2021

Classification (1st year): 19 (out of 20)
Graph-Based Entity-Oriented Search
Thesis statement: A graph-based joint representation of unstructured and structured data has the potential to unlock novel ranking strategies, that are, in turn, able to support the generalization of entity-oriented search tasks and to improve overall retrieval effectiveness by incorporating explicit and implicit information derived from the relations between text found in corpora and entities found in knowledge bases.
During the course of this doctoral work, two main systems were developed: ANT, an entity-oriented search engine for the University of Porto (prototype); and Army ANT, a workbench for innovation in entity-oriented search.

Doctoral Program in Computer Science (MAP-i)

https://ant.fe.up.pt/ https://github.com/feup-infolab/ant (private)

https://github.com/feup-infolab/army-ant https://github.com/feup-infolab/army-ant-install https://hub.docker.com/r/jldevezas/army-ant

2003-2010 Master in Informatics and Computing Engineering U.Porto Classification: 14 (out of 20)

Link Ecosystem of the Portuguese Blogosphere

This thesis explored the blogosphere as an ecosystem of blogs comprised not only of a set of posts and their individual content, but also of the interactions established through the hyperlinks connecting the posts. Based on this study, we discovered a correlation between the blog's centrality and the length and number of posts.

Thesis Classification: 18 (out of 20)

Technological Specialization in Informatics 2000-2003 Classification: 18 (out of 20)

Colégio de Gaia

Tallinn. Estonia

U.Porto, UA, UMinho

Experience

2022-Now

HEX DATUM OÜ Founder CEO

Hex Datum OÜ was founded as a solopreneur and Estonian e-residency company. We sell software-as-a-service (SaaS), as well as consulting services in the area of intelligent information systems. Our flagship product is a crypto analytics web application called Indicator Dash (indicatordash.app). Key responsibilities during the development of this service included:

Data Engineering:

- Python-based ETL;
- PostgreSQL and JanusGraph DBA, and relational and graph data modeling;
- Redis caching layer to ensure real-time usability.

Data Science:

- Design and implementation of technical, fundamental, and sentiment financial indicators;
- Data analysis and visualization for small blog-based studies;
- ML prototyping in Jupyter (time series forecasting using transformers and LSTMs).

Search Engineering:

- Entity search for people, organizations and products;
- Entity-oriented news search and ranking;
- Sentiment-based news filtering.

Full-Stack Development:

- Python-based stack built on top of Flask;
- JavaScript-based stack built on top of Vue.js and Tachyons.

Data Presentation:

- Financial charting (candlestick/OHLC);
- Knowledge graph visualization and exploration UI.

2021–2022 CAREER BREAK

Career Transition

After completing the PhD, I decided to invest in building my own business. I started developing a crypto analytics prototype web application, which eventually led to the opening of a company, Hex Datum OÜ, as an e-resident of Estonia.

2016–2021 **CSIG, INESC TEC**

Researcher / PhD Student

MAP-i PhD student and FCT grant holder, working on entity-oriented search. Former FourEyes researcher with a focus on studying information consumption on social media. Provided support to the research group through data engineering, by developing the infrastructure for Twitter data collection and analysis, as well as news collection and archiving.

2015–2021 INFORMATION SYSTEMS LABORATORY, UNIVERSITY OF PORTO

Researcher / PhD Student MAP-i PhD student and FCT grant holder, working on entity-oriented search.

Research and development of the ANT entity-oriented search engine. Particularly focused on query analysis methodologies, as well as the construction of specialized indexes and ontologies for efficient and effective questionanswering through contextual widgets. The produced system was also aimed at providing a platform able to support the academic community with the experimentation of search engines backed by linked data.

2017–2018 **FACULTY OF ENGINEERING, UNIVERSITY OF PORTO** Porto, Portugal *Invited Assistant* Lectured one semester on social network analysis, for the Information Manage-

Lectured one semester on social network analysis, for the Information Management in Social Networks course, at the Master in Information Science, and one semester on relational and document databases, for the Databases course, at the Master in Informatics and Computing Engineering.

Porto, Portugal

Porto, Portugal

Porto, Portugal

2014–2015

Software Engineer / Data Scientist Data Science:

- Collection and analysis of several types of data, mainly originating from the web, particularly from social media.
- Highly proficient in R, Python and SQL (from data collection to reporting), including topic tracking, outlier and anomaly detection, statistical smoothing, regression, classification, clustering and network analysis.
- Skilled in dynamic reporting.
- Developed several data visualization widgets.
- Analyzed social networks and other graphs.

Software Engineering:

INTERRELATE

- Software design of web applications and analytics pipelines.
- Technology evaluation and selection.
- Supporting the team with new technologies or languages.
- INTERRELATE Backoffice:
 - Devising strategies for automatic and periodical data collection and information extraction from unstructured user-generated content;
 - Systems integration, including team coordination (4 person project).

Programming:

- Worked in nearly every company project, using various different languages.
- Developed several web applications and web services.
- INTERRELATE Backoffice:
 - Web crawler development;
 - Development of data visualization widgets for internal statistics;
 - Front-end design and development.

Sysadmin:

- Manage the company servers, including security, backups and product deploys.
- Maintain firewall rules and SMTP server.
- DNS configurations (including SPF, for SPAM control).
- Nginx configurations.
- Company services maintenance (internal and external).
- Log monitoring and bug reporting.
- Real-time e-mail log error monitoring for critical systems and web service mobile monitoring for downtime or connection issues.
- Support Ticket System configuration.
- PayPal payment services.

2012–2013 SAPO LABS, UNIVERSITY OF PORTO

Researcher / Developer

Porto, Portugal

Porto, Portugal

Research and development of music discovery and recommendation techniques supported on a graph-based neighborhood approach, as well as on latent factor models based on matrix factorization algorithms. Development of methodologies to combine knowledge from different dimensions, including not only user profile, content and context features, but also social features, such as community structure. Related tasks required work on data mining, multimodal network analysis and community detection, and information retrieval.

2011-2012

012 CRACS, INESC TEC

Researcher / Developer

Enhancement, adaptation and implementation of state of the art algorithms for network analysis, data mining and data visualization, in the context of intelligent information systems for cyber journalism, as part of the Breadcrumbs Project (UTA-Est/MAI/0007/2009).

2010-2011	SAPO LABS, UNIVERSITY OF PORTO
-----------	--------------------------------

Porto, Portugal

Researcher / Developer Work in the areas of data mining, network analysis, community detection, information retrieval, real-time data analytics and data visualization.

2002-2003 NPF PORTUGUESE FREEBSD GROUP - npf.pt.freebsd.org Coimbra, Portugal Documentation / Programming Translation of the official FreeBSD documentation to portuguese, creation of how-to guides in portuguese, porting open source applications to FreeBSD and programming and content management of a hints widget.

Awards

2012 **Certificate of Merit** The 2012 IAENG International Conference on Data Mining and Applications Awarded to the best papers published in the conference.

Communication Skills

2020	Demo European Conference on Information Retrieval Prepared a video demonstration of the Army ANT system, which was exhibited during the online event, with a slot reserved for questions.
2020	Doctoral ConsortiumEuropean Conference on Information RetrievalParticipated in the doctoral consortium, presenting graph-based entity-orientedsearch as a unified framework in information retrieval.
2020	Lecture Software Systems Architecture course, at the Master in Informatics and Computing Engineering Described ANT search engine's system architecture, focusing on the requirements of an entity-oriented search engine.
2019	PresentationSymposium on Languages, Applications and TechnologiesPresented the graph-of-entityrepresentation and retrieval model for entity-oriented search over combined data.
2019	Invited Speaker Creative CoLAB 2019 Introduced several concepts about hypergraphs, arguing about their generality, and presented an approach to mapping terms and entities with hypergraphs, which could be used to universally solve multiple entity-oriented search tasks.
2018	Lecture Information Description, Storage and Retrieval course, at the Master in Informatics and Computing Engineering Described ANT search engine's system architecture, introducing several technical aspects of the building of an entity-oriented search engine.
2018	WorkshopIEEE Student and Young Professional CongressPresented ANT, focusing on its query understanding algorithm, based on the Score Hypergraph, as well as and Army ANT, showing how it an be used to research entity-oriented search.
2017	PresentationSymposium on Languages, Applications and TechnologiesPresented a complete pipeline, from data acquisition, passing through information extraction and the automatic construction of knowledge base, to an information retrieval implementation.
2013	Presentation DEI, FEUP Did a short presentation on the Juggle project for the department, in the context of DEI Talks.

2013	Presentation	ect, including the developed	SAPO
	Presented the Juggle projection algorithms, for the	SAPO technical team.	I graph-based recommen-
2012	Presentation Internation Presented an interactive r istic research and knowle	nal Conference on Knowledge Dis news clips visualization tool dge discovery.	covery and Information Retrieval for applications in journal-
2012	Presentation	IAENG International Conference	on Data Mining and Applications
	Presented a methodology	, based on community det	ection in multidimensional
	networks, for the creation	of news context from a foll	<sonomy clipping.<="" of="" td="" web=""></sonomy>
2011	Poster	International AAAI Conference	e on Weblogs and Social Media
	Presented research work	on using the h-index to est	imate blog authority.
2010	Presentation Presented Ciclope project part of the inaugural presented	t for the Portugal Telecom? entation of Laboratório SAF	SAPO Labs, University of Porto s CEO and the media, as PO/U.Porto.
2010	Poster	KDD Worl	kshop on Social Media Analytics
	Presented the blog popul	arity study conducted durin	g my Masters degree.

Interests

Professional: search engines, data analysis, web development, blockchain, DeFi, algorithms and data structures, data visualization **Personal:** investment, crypto, games, music, tv, movies

More information: https://josedevezas.com/

Collaborations

2020	Member of the program committee Graphs and More Complex Structures for Learning and Reasoning, Workshop At AAAI 2021 https://sites.google.com/view/gclr2021/
2019–2020	Member of the program committee European Conference on Information Retrieval 2020 and 2021 https://www.ecir2020.org/, https://www.ecir2021.eu/
2017–2019	Member of the program committee Complex Networks 2017, 2018 and 2019 http://complexnetworks.org
2019/2020	Master's thesis co-supervision Building a domain-specific search engine that explores football-related search patterns https://hdl.handle.net/10216/128526
2017/2018	Participation in TREC 2018 with the University of Alicante FEUP at TREC 2018 Common Core Track – Reranking for Diversity using Hypergraph-of-Entity and Document Profiling https://trec.nist.gov/pubs/trec27/papers/FEUP-CC.pdf
2016/2017	Master's thesis co-supervision Named entity extraction from Portuguese web text http://hdl.handle.net/10216/106094
2015/2016	Master's thesis co-supervision Exploring the Sea: Heterogeneous Geo-Referenced Data Repository http://hdl.handle.net/10216/85612

Publications

Journal Articles

- A Review of Graph-Based Models for Entity-Oriented Search José Devezas and Sérgio Nunes SN Computer Science 2.6 (2021). 2021
- Characterizing the hypergraph-of-entity and the structural impact of its extensions José Devezas and Sérgio Nunes Applied Network Science - Special Issue of the 8th International Conference on Complex Networks and Their

Applications 5.1 (2020) p. 79. 2020, DOI: 10.1007/s41109-020-00320-z

Hypergraph-of-entity: A unified representation model for the retrieval of text and knowledge José Devezas and Sérgio Nunes

Open Computer Science 9.1 (June 2019) pp. 103–127. 2019, DOI: 10.1515/comp-2019-0006

Graph-Based Entity-Oriented Search: Imitating the Human Process of Seeking and Cross Referencing Information

José Devezas and Sérgio Nunes

ERCIM News. Special Issue: Digital Humanities 111 (Oct. 2017) pp. 13–14. 2017, URL: https://ercim-news.ercim. eu/en111/special/graph-based-entity-oriented-search-imitating-the-human-process-of-seeking-and-crossreferencing-information

The community structure of a multidimensional network of news clips

José Luís Devezas and Álvaro Reis Figueira Int. J. Web Based Communities 9.3 (2013) pp. 411–429. 2013, DOI: 10.1504/IJWBC.2013.054911

Finding Language-Independent Contextual Supernodes on Coreference Networks José Devezas and Álvaro Figueira

IAENG International Journal of Computer Science 39.2 (2012) pp. 200–207. 2012, URL: http://www.iaeng.org/ IJCS/issues_v39/issue_2/IJCS_39_2_07.pdf

Conference Articles

Federated Search Using Query Log Evidence

João Damas, José Devezas, and Sérgio Nunes

Progress in Artificial Intelligence - 21st EPIA Conference on Artificial Intelligence, EPIA 2022, Lisbon, Portugal, August 31 - September 2, 2022, Proceedings, 2022, DOI: 10.1007/978-3-031-16474-3_64

Graph-Based Entity-Oriented Search: A Unified Framework in Information Retrieval José Devezas

Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II, 2020, DOI: 10.1007/978-3-030-45442-5_78

Army ANT: A Workbench for Innovation in Entity-Oriented Search

José Luís Devezas and Sérgio Nunes

Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II, 2020, DOI: 10.1007/978-3-030-45442-5_56

Characterizing the Hypergraph-of-Entity Representation Model

José Devezas and Sérgio Nunes

Complex Networks and Their Applications VIII - Volume 2 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019, Lisbon, Portugal, December 10-12, 2019, 2019, DOI: 10.1007/978-3-030-36683-4_1

Graph-of-Entity: A Model for Combined Data Representation and Retrieval

José Devezas and Sérgio Nunes

Proceedings of the 8th Symposium on Languages, Applications and Technologies (SLATE 2019), 2019, Vila do Conde, Portugal, DOI: 10.4230/OASIcs.SLATE.2019.1

Social Media and Information Consumption Diversity

José Devezas and Sérgio Nunes

Proceedings of the Second International Workshop on Recent Trends in News Information Retrieval co-located with 40th European Conference on Information Retrieval (ECIR 2018), Grenoble, France, March 26, 2018, 2018, URL: http://ceur-ws.org/Vol-2079/paper5.pdf

FEUP at TREC 2018 Common Core Track - Reranking for Diversity using Hypergraph-of-Entity and Document Profiling

José Luís Devezas, Sérgio Nunes, Antonio Guillén, Yoan Gutiérrez, and Rafael Muñoz Proceedings of the Twenty-Seventh Text REtrieval Conference, TREC 2018, Gaithersburg, Maryland, USA, November 14-16, 2018, 2018, URL: https://trec.nist.gov/pubs/trec27/papers/FEUP-CC.pdf

Information Extraction for Event Ranking

José Devezas and Sérgio Nunes

6th Symposium on Languages, Applications and Technologies (SLATE 2017), 2017, Dagstuhl, Germany, DOI: 10.4230/OASIcs.SLATE.2017.18

FEUP at TREC 2017 OpenSearch Track Graph-Based Models for Entity-Oriented

José Luís Devezas, Carla Teixeira Lopes, and Sérgio Nunes Proceedings of The Twenty-Sixth Text REtrieval Conference, TREC 2017, Gaithersburg, Maryland, USA, November 15-17, 2017, 2017, URL: https://trec.nist.gov/pubs/trec26/papers/FEUP-O.pdf

Benchmarking Named Entity Recognition Tools for Portuguese

André Pires, José Devezas, and Sérgio Nunes

Proceedings of Simpósio de Informática (INForum 2017), 2017, Aveiro, Portugal

Index-Based Semantic Tagging for Efficient Query Interpretation

José Luís Devezas and Sérgio Nunes

Experimental IR Meets Multilinguality, Multimodality, and Interaction - 7th International Conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016, Proceedings, 2016, DOI: 10.1007/978-3-319-44564-9_17

Exploring a Large News Collection Using Visualization Tools

Tiago Devezas, José Devezas, and Sérgio Nunes

Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016, 2016, URL: http://ceur-ws.org/Vol-1568/paper9.pdf

Large-scale crossmedia retrieval for playlist generation and song discovery

Filipe Coelho, José Luís Devezas, and Cristina Ribeiro

Open research Areas in Information Retrieval, OAIR '13, Lisbon, Portugal, May 15-17, 2013, 2013, URL: https://dl.acm.org/doi/10.5555/2491748.2491764

Using the overlapping community structure of a network of tags to improve text clustering Nuno Cravino, José Luís Devezas, and Álvaro Figueira

23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012, 2012, DOI: 10.1145/2309996.2310036

Creating News Context From a Folksonomy of Web Clipping

José Devezas, Henrique Alves, and Álvaro Figueira

Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2012 (IMECS 2012), 2012, Hong Kong, URL: http://cracs.fc.up.pt/sites/default/ files/c2012_arf_IAENG.pdf

Studying a Personality Coreference Network in a News Stories Photo Collection

José Luís Devezas, Filipe Coelho, Sérgio Nunes, and Cristina Ribeiro

Advances in Information Retrieval - 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings, 2012, DOI: 10.1007/978-3-642-28997-2_47

Interactive Visualization of a News Clips Network: A Journalistic Research and Knowledge Discovery Tool

José Devezas and Álvaro Figueira

Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2012), 2011, Barcelona, Spain, URL: https://scitepress.org/papers/2012/41087/41087.pdf

Using the H-Index to Estimate Blog Authority

José Luís Devezas, Sérgio Nunes, and Cristina Ribeiro

Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011, 2011, URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2829

FEUP at TREC 2010 Blog Track: Using h-index for blog ranking

José Luís Devezas, Sérgio Nunes, and Cristina Ribeiro Proceedings of The Nineteenth Text REtrieval Conference, TREC 2010, Gaithersburg, Maryland, USA, November 16-19, 2010, 2010, URL: http://trec.nist.gov/pubs/trec19/papers/labs-sapo-up.blog.pdf

Studying blog features over link popularity

José Luís Devezas, Cristina Ribeiro, and Sérgio Nunes

Proceedings of the 3rd Workshop on Social Network Mining and Analysis, SNAKDD 2009, Paris, France, June 28, 2009, 2010, DOI: 10.1145/1964858.1964863

Technical Reports

Army ANT: A Workbench for Innovation in Entity-Oriented Search - External Option: Scientific Activities – TREC Open Search

José Devezas

Research rep., 2017, URL: https://hdl.handle.net/10216/110181

Auditing Open Access Repositories - Free Option: Supervised Study – Digital Archives and Libraries

José Devezas Research rep., 2017, URL: https://hdl.handle.net/10216/104152

Entity-Oriented Search - FEUP InfoLab Report

José Devezas Tech. rep., 2016

An Overview of the Graph Database Paradigm - Breadcrumbs Project José Devezas

Tech. rep., 2011

Overlapping Community Detection - Laboratório SAPO/U.Porto Report José Devezas, Sérgio Nunes, and Cristina Ribeiro *Tech. rep., 2011*

Book Chapters

Creating and Analysing a Social Network Built From Clips of Online News Álvaro Figueira, José Devezas, Nuno Cravino, and Luis-Francisco Revilla Information Systems and Technology for Organizations in a Networked Society, *chap. 5, pp. 67–86, IGI Global,* 2012, URL: https://www.igi-global.com/chapter/creating-analysing-social-network-built/76532

Theses

Graph-Based Entity-Oriented Search José Devezas PhD thesis 2021, URL: http://josedevezas.com/publications/mapi1617-jose_devezas-thesis-digital_version.pdf Link Ecosystem of the Portuguese Blogosphere José Devezas

Master's thesis 2010, URL: http://hdl.handle.net/10216/58922

Datasets

Simple English Wikipedia Link Graph with Clickstream Transitions 2018-12 [dataset] José Devezas and Sérgio Nunes

INESC TEC research data repository, 2019, DOI: 10.25747/83vk-zt74

Selected Projects

Here I present some of the most representative projects of my research career, where I developed multiple competences in the broad areas of information retrieval, network science, and machine learning. This serves the purpose of illustrating not only the challenges I tackled, but also the feasibility, from start to finish, of all the projects I participated in, where a working prototype was always delivered, as a live demonstration of the carried research.

2021–2023 Indicator Dash Hex Datum OÜ Indicator Dash is a crypto analysis service designed to help you "do your own research". It provides information about multiple cryptocurrencies, including ranked news and tweets, as well as basic price charting and indicators, including custom metrics, which are developed in-house.

Indicators are translated into normalized signals that can then be used to personalize the ranking of currency pairs, or for illustrating the bullish, rangebound, or bearish nature of the market, in the context of that particular pair. That is, it attempts to identify anomalous price behavior, so that users can better identify buying and selling opportunities.

The user is offered state-of-the-art decision support technology, based on news and entity search, sentiment analysis, interactive knowledge graph visualization and exploration, community detection with related, news-driven, automatic text summarization, as well as price forecasting based on machine learning models.

Internally, several studies were ran, using Jupyter Notebooks, in order to prototype code, that later on was integrated into the production system (e.g., price forecasting using different neural network architectures, namely LSTMs and transformers), or to develop ad hoc analyzes with valuable insights, some of which were posted in our blog (e.g. best time to trade the top 10 cryptocurrencies).

While the code is private, the GitHub for the company is available at https: //github.com/hexdatum. Upon timely request, a demo of this application can be provided.





Buy ranking	Threshold	-0.75	Sell ranking	Threshold
Rank	Pair	Score	Rank	
1	LUNA/TRY	-0.1099	1	PIVX/BTC
2	BUSD/VAI	-0.1077	2	NEBL/BTC
3	TRX/XRP	-0.0849	3	VIB/BTC
4	ETH/BIDR	-0.0731	4	PROS/ETH
5	YFI/EUR	-0.0725	5	QLC/BTC
6	KLAY/BTC	-0.0672	6	VIB/ETH
7	XVS/BNB	-0.0664	7	OAX/BTC
8	HOD/USDT	-0.0656	8	BETH/ETH
9	AVAX/EUR	-0.0649	9	AST/BTC
10	ZIL/USDT	-0.0625	10	ATOM/CRO
11	STORJ/BUSD	-0.0625	11	AMB/BTC
12	ETH/DAI	-0.0614	12	GLM/BTC
13	KSM/BNB	-0.0614	13	PERL/BTC
14	RUNE/GBP	-0.0608	14	HARD/USDT
15	UMA/USDT	-0.0601	15	BUSD/BRL
16	DOT/EUR	-0.0599	16	GODS/USDT
17	RGT/USD	-0.0599	17	LINK/BNB
18	ETC/EUR	-0.0595	18	FIRO/USDT
19	ETC/GBP	-0.0591	19	PERL/USDT
20	DOGE/EUR	-0.0581	20	SSV/ETH

Community 265456 (25,006 entities)

- In August, the <u>CME Group</u> launched two types of derivative products, euro-denominated <u>bitcoin</u>, and <u>ether futures</u> contracts, designed to help institutional investors mitigate risks associated with the volatility of cryptocurrencies, <u>McCourt</u> said at the time.
- Adding fire to the already politically charged debate, <u>Porter</u> suggested the \$10,000 due to qualifying persons could be used to buy <u>Bitcoin</u>, which is then locked in a smart contract that enables access after ten years.
- A recent example includes the <u>Royal Museum of Fine</u> Arts Antwerp's fractionalization and security <u>token</u> offering on the <u>Polygon blockchain</u> of <u>James Ensor's</u> 1924 painting "Carnaval de Binche."
- Most recently, two members of the <u>house of</u> <u>Representatives</u>, Congresswoman Norma J. <u>Torres</u> of California, and Congressman Rick <u>Crawford</u> of Arkansas

Members by EntityRank

People (7,775)



- Center an open-source technology project launched by founding members of Circle and Colibbase – raised \$11 million in a seed round led by <u>Thrive Capital</u> and <u>Founders Fund</u> to build a search index.
- That specific degree of support includes proposals for FedAccounts that initially appeared in 2020, as well as moves to expand the role of the US Postal Service in payments and accounts, which Rep. Ayanna Pressley (D-MA) praised during today's hearing.
- After withdrawing its motion to hire back its ex-CFO <u>Rod Bolger</u> at \$92,000 a month, current CFO <u>Chris</u> <u>Ferraro</u> has come <u>forward</u> to address the reports that have been going around that the <u>company</u> would run out of money in October.



2017–2021 Army ANT

Laboratório de Sistemas de Informação/U.Porto

Army ANT is an information retrieval research framework that supports experimentation with classical approaches, while also providing a high-level abstraction that motivates the implementation of innovative approaches in a shared evaluation environment. It facilitates the use of combined data, and it provides an environment for testing and evaluating multiple retrieval tasks, supporting keyword or entity queries, as well as documents, entities, or even terms, as the rankable results.

This software contains the majority of the code I developed throughout my PhD, providing iterators for well-known test collection, as well as evaluators for several common retrieval tasks from the TREC and INEX evaluation forums. Several experimental search models are also available in Army ANT, along with classical approaches based on Apache Lucene.

The code is available at https://github.com/feup-infolab/army-ant, and several Docker images are also available at https://hub.docker.com/r/jldevezas/ army-ant, which can be installed using Docker Compose based on the code at https://github.com/feup-infolab/army-ant-install.

Configure output						o c	SV 🔵 LaTeX	
Toggle i	all Select Favo	orite						
GMAP		MAP		Mac	ro Avg F0_5		Macro Avg	F1
Macro A	vg F2	🗹 Macro Av	g Prec	🗹 Mac	ro Avg Rec		Micro Avg F	0_5
Micro Av	/g F1	Micro Av	g F2	Micr	o Avg Prec		Micro Avg F	Rec
VDCG@	10	NDCG@1	.00	NDC	G@1000	\checkmark	P@10	
P@100		P@1000						
Decimals	^ ~	Columns	🖊 Type 🔽 F	Parameters	Location			
Туре	Parameters		MAP Ma	cro Avg Pred	: Macro Avg Re	c NDCG@	10 P@10	
hgoe	(I=2, r=10)		0.0941 0.10)23	0.0231	0.0207	0.0900	
hgoe	(I=2, r=100)		0.1017 0.10)52	0.0244	0.0226	0.1200	
hgoe	(I=2, r=1000)	0.1501 0.17	38	0.0408	0.0351	0.1500	
hgoe	(I=3, r=10)		0.0814 0.10	88	0.0239	0.0160	0.0900	
hgoe	(I=3, r=100)		0.1355 0.13	318	0.0290	0.0335	0.1500	
hgoe	(l=3, r=1000)	0.1377 0.14	106	0.0341	0.0255	0.1300	
hgoe	(I=4, r=10)		0.1221 0.12	245	0.0271	0.0253	0.1400	
hgoe	(l=4, r=100)		0.1236 0.15	646	0.0377	0.0303	0.1600	
						[Cancel Q	Download

2015–2021 **ANT**

ANT Laboratório de Sistemas de Informação/U.Porto ANT is an entity-oriented search engine available at http://ant.fe.up.pt, that indexes data from the SIGARRA Information System. On one side, it aims at improving search in U.Porto. On the other side, it aims at providing a platform to educate students and to support future research in the area.

This research project motivated my doctoral proposal, providing the ideal platform for the integration of previous areas of interest, particularly information retrieval and network science, but also recommendation and personalization. Working in ANT led me to identify several challenges in the area of entity-oriented search and to devise an innovative and focused contribution.

In a similar fashion to Google and other modern search engines, this enabled our system to answer user questions more directly, well beyond the traditional keyword-based matching. Our proposed integrated solution for segmentation and annotation of queries is illustrated in the following figure of the development front-end.



2012–2013 Juggle

Laboratório SAPO/U.Porto

Juggle project aimed at improving music discovery based on a hybrid largescale recommender system, capable of handling and combining different types of data, namely text and audio content, context from elements such as tags or location, and collaborative information from user profiles.

We tackled the challenges of multimodality and large-scale by developing a graph-based recommender system, supported on Neo4j, a popular and robust graph database that facilitated the modeling of content, context and collaborative information as nodes and edges in a graph. One of the biggest challenges was the translation of audio content to relationships in a graph, specifically the comparison of the audio features of a million songs with each other, which we solved by using an approximate search algorithm from image retrieval.

Our recommendation algorithm was mainly supported on neighborhood methods for collaborative filtering, but we also used metrics from text retrieval to boost the relevance of tags in the long tail, while not completely disregarding tag popularity, in order to offer a playlist that better potentiated the discovery of music.



In Juggle Mobile we presented the users with the ability to create an account and fill their taste profiles either based on our random artist rating system, or by importing their existing music information from Facebook or Last.fm. All the data from these different sources was mixed together based on our weighting model and used to provide recommendations to the user or to a group of nearby users.

Our experiments were based on a linear algebra approach, where, instead of a graph, we used a user-items matrix, applying singular value decomposition to build a latent factor model that provided the support for individual and group recommendations. For groups, we proposed a rating aggregation method that ensured an equal chance for every group member to have a relevant influence in the recommendations outcome.



2011-2012

CRACS/INESC TEC

Breadcrumbs

As a Breadcrumbs researcher, I was able to make contributions on several different areas. I implemented a language-independent named entity recognition system based on DBpedia entity lists. This system enabled the identification of three different types of entities - people, places and dates - tied to three of the five dimensions (the Five-Ws) of journalism: who, where and when. Using this data, a multidimensional entity coreference network was built, connecting news clips that cited the same entity. Next, I implemented the community detection methodologies for multidimensional networks proposed by Tang et al. This included the dimension integration strategies proposed by the authors, based on their unified view of four traditional community detection methodologies. These algorithms were also implemented in the system, along with the Louvain method, one of the state of the art algorithms for community detection.

Next, two visualization tools were developed to display and explore the acquired data. The first was analogous to a map, where communities were visualized as countries resulting of the aggregation of a node population. The second enabled the exploration of the multidimensional network based on the three identified dimensions: who, where and when. Some simple chart visualizations were created to display statistics about the top user and system tags and entities.

We used a topic model, based on Latent Dirichlet Allocation, to suggest titles for each collection of news clips; a simplistic event detection system was also created, in order to find relevant peaks of activity in a time series of entity frequencies. Some other trivial systems, such as an administration panel, capable of scheduling tasks, and a widget dashboard were also implemented.

These algorithms were all developed using a web services architecture, communicating using either XML or JSON. Several scientific papers were published as the results of the described research. Below are some screenshots of the Breadcrumbs modules I contributed to in some way.







2010–2011 Ciclope

http://blogs.blogs.sapo.pt/133502.html

2010-10-12 15:14:20

One of my first projects was Ciclope, a real time data visualization project aimed at gathering information from SAPO Blogs clickstream and displaying it in a useful way, allowing the blog owner to have an understanding of how the traffic flow of his or her blog behaved.

Among other widgets, we developed two main visualizations: a real time bar chart that displayed the number of visits per second along with a table showing traffic origin and destination; and a custom flow tree to visualize and quantify aggregated traffic sources for a given blog.



http://blogs.blogs.sapo.pt/137696.html

