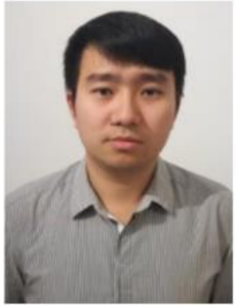
Knowledge Graph – Enrich the Results in Search Engine and Recommender System
DBKDA - InfoSys 2021

Yan WANG
Capgemini Engineering, Direction of Research and Innovation (DRI)
yan.wang2@altran.com

altran
Part of Capgemini

IARIA

Dr. Yan WANG



Current	Research Project Manager	DRI, Capgemini Engineering
2020	Computer Science Postdocs	LIP6, Sorbonne University
2019	Computer Science Postdocs	CRI, Paris 1 Panthéon-Sorbonne University
2018	Enterprise Engineering PhD degree	IMS, University of Bordeaux
2014	Enterprise Engineering Master	IMS, University of Bordeaux
2014	Software Engineering Master	Harbin Institute of Technology
2012	Computer Science Bachelor	Harbin Institute of Technology

Research interests :

- Discrete Event Modeling and Simulation
- Process Mining
- Fuzzy Logic
- Constraint Programming
- Reinforcement Learning
- NLP
- Knowledge Graph
- Recommender System.

Research and Development Project TNT

TNT (Talent Needs Trends) is a research and development project of the program Future of Engineering. The objective of this project is to propose a competence management system advanced and adapted to Capgemini. The purpose is to improve the synergy between skills, resources and customers, simplify the process of HR-Analytics.

TNT is launched from 2014 with a lot of propositions and development tools. This year, we focus on the matching tool with three parts – search engine, recommender system and constraint solver.

Context of Project TNT

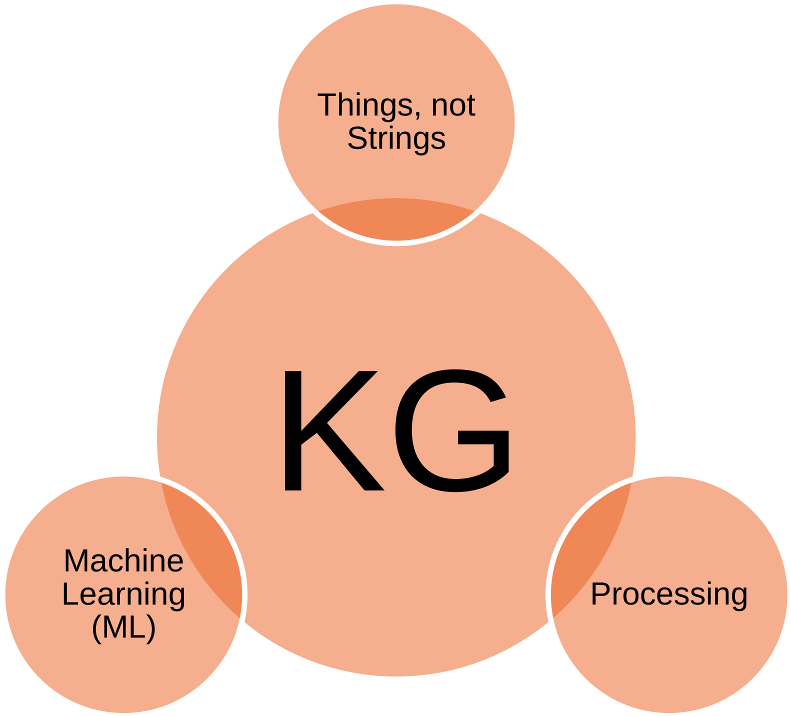
- The rapid evolution of the competences of the candidates
  - The continuing evolution of candidate experience and technology

- The rapid evolution of the requirements of the clients

- HR adaptation and treatment time is getting longer and longer

Knowledge Graph (KG)

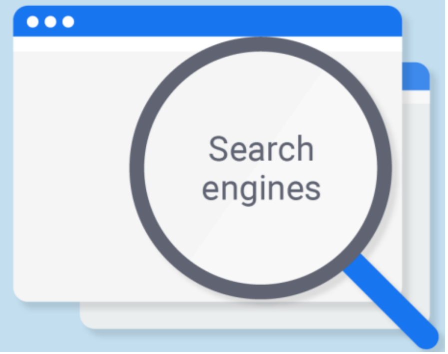
Knowledge Graph (KG) has been proposed to discover the relation information with the property of powerful language understanding and rapid data analysis. It is first proposed in 2012 by Google, a theory of semantic structure combining applied mathematics, computer graphics, information visualization and machine learning. Knowledge Graph is constructed based on “Entity-Relation-Entity” with the associated property on entity.

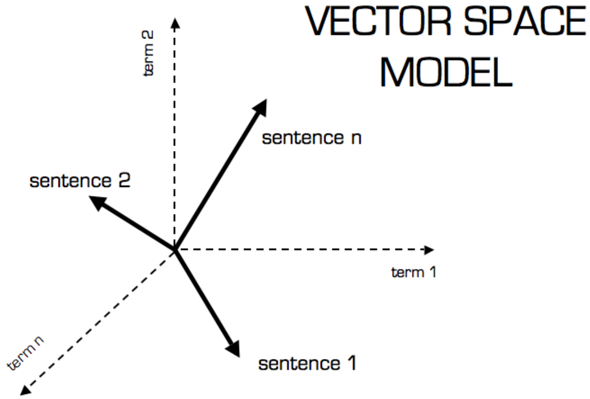


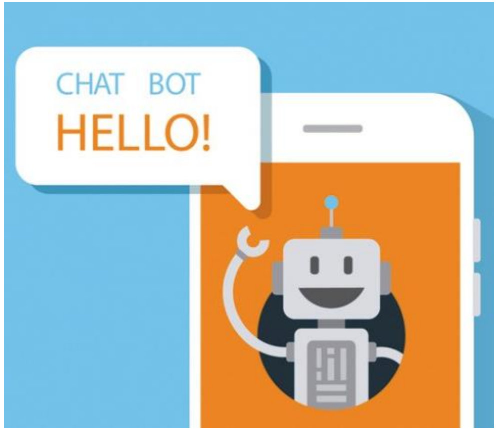
1.KG in ML

Typical Knowledge Graph

Application Scenarios of the Knowledge Graph







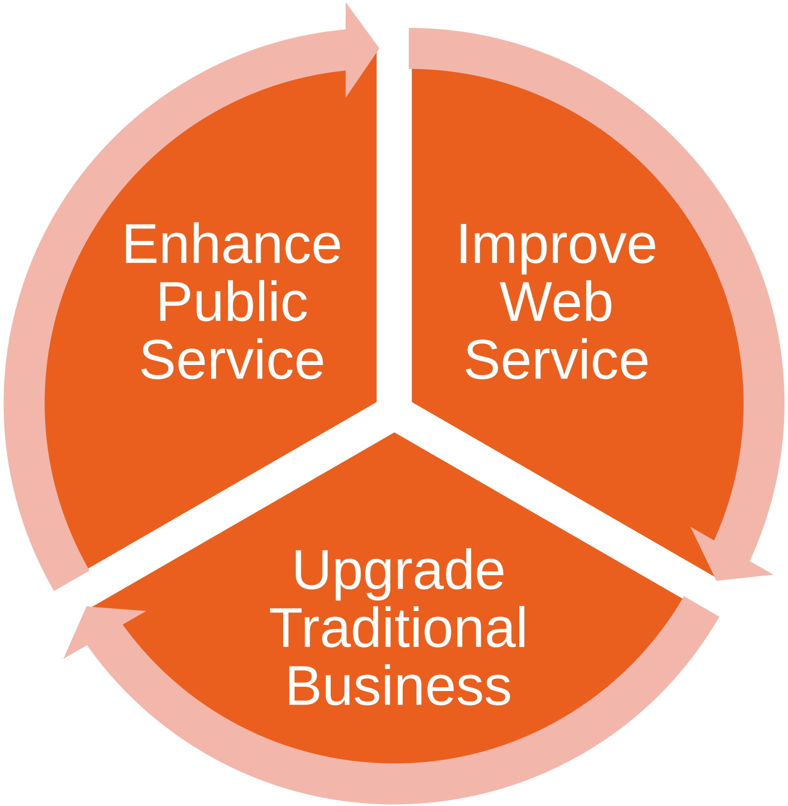
Search Engine in history

Recommender System in History

1998 Amazon item-to-item recommendation
2004-Now Special sessions in recommender system in several important conferences & journals:
AI Communications ; IEEE Intelligent Systems; International Journal of Electronic Commerce; International Journal of Computer Science and Applications; ACM Transactions on Computer-Human Interaction; ACM Transactions on Information Systems
2007 First ACM RecSys conference
2008 Netflix online services (& innovative HMI)
2008-09 Netflix RS prize
2010-Now RS become essential : YouTube, Netflix, Tripadvisor, Last.fm, IMDb, etc...

Why Knowledge Graph?

With the help of KG, users can get a more accurate recommendation as well as the explanations for recommended items. (Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, and Q. He. A survey on knowledge graph-based recommender systems. IEEE Transactions on Knowledge and Data Engineering. 2020)



Typical Knowledge Graph

K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, “Freebase: a collaboratively created graph database for structuring human knowledge,” in Proceedings of the 2008 ACM SIGMOD international conference on Management of data. AcM, 2008, pp. 1247–1250.

J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer et al., “Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia,” Semantic Web, vol. 6, no. 2, pp. 167–195, 2015.

F. M. Suchanek, G. Kasneci, and G. Weikum, “Yago: a core of semantic knowledge,” in Proceedings of the 16th international conference on World Wide Web. ACM, 2007, pp. 697–706.

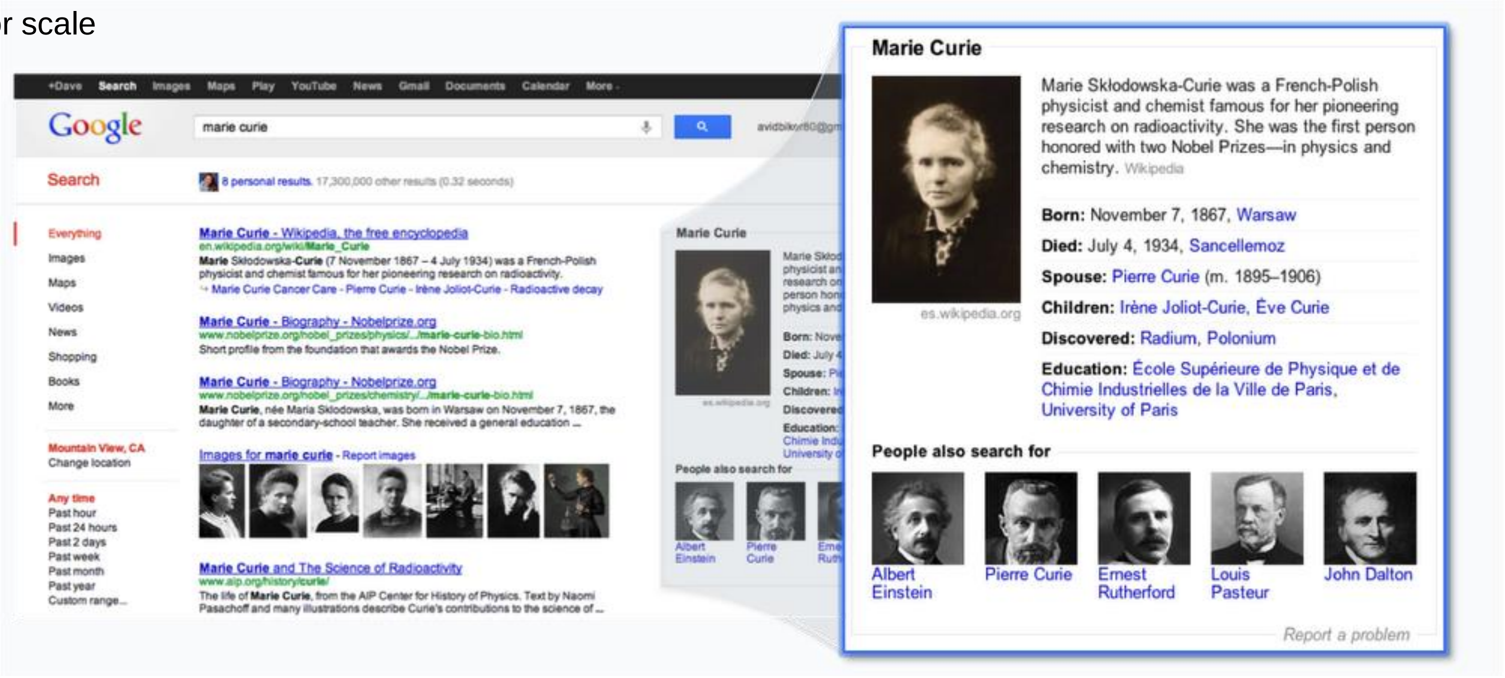
A. Singhal, “Introducing the knowledge graph: things, not strings,” 2012, https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.

2. Things, not Strings

Knowledge Representation
Knowledge Modeling

Knowledge Representation

- Scope of the knowledge
- Compatible for machine
- Structural for scale



altran Part of Capgemini

Knowledge Modeling

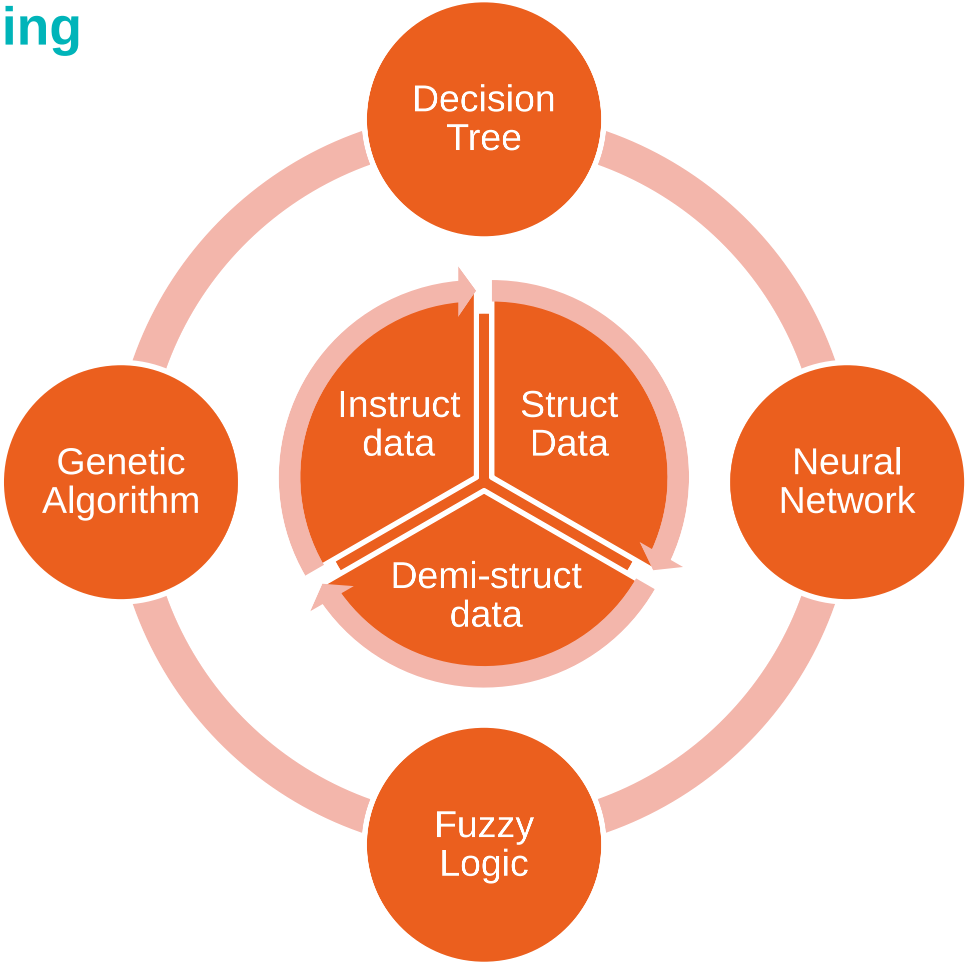
RDF = Resource Description Framework
Purpose : to provide a structure for describing identified things

OWL = Web Ontology Language
Purpose : to develop ontologies that are compatible with the World Wide Web

3. Processing

Knowledge Mining
Knowledge Storage
Knowledge Query
Knowledge Analytics

Knowledge Mining



Tools for Knowledge Storage

OrientDB®

APACHE Spark™

rdf4j

JanusGraph

Dgraph
The GraphQL Database

HyperGraphDB

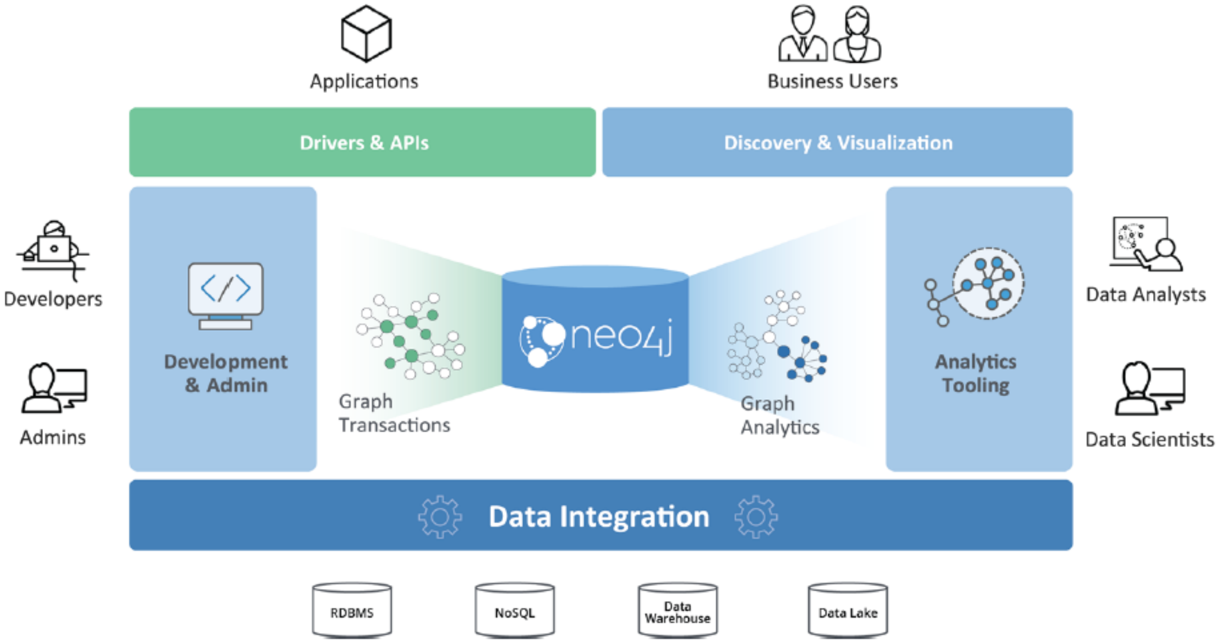
HugeGraph
a fast-speed and high-scalable graph database

neo4j

altran Part of Capgemini

Neo4j

The Neo4j Graph Platform is an example of a tightly integrated graph database and algorithm-centric processing, optimized for graphs. It is popular for building graphbased applications and includes a graph algorithms library tuned for its native graph database.



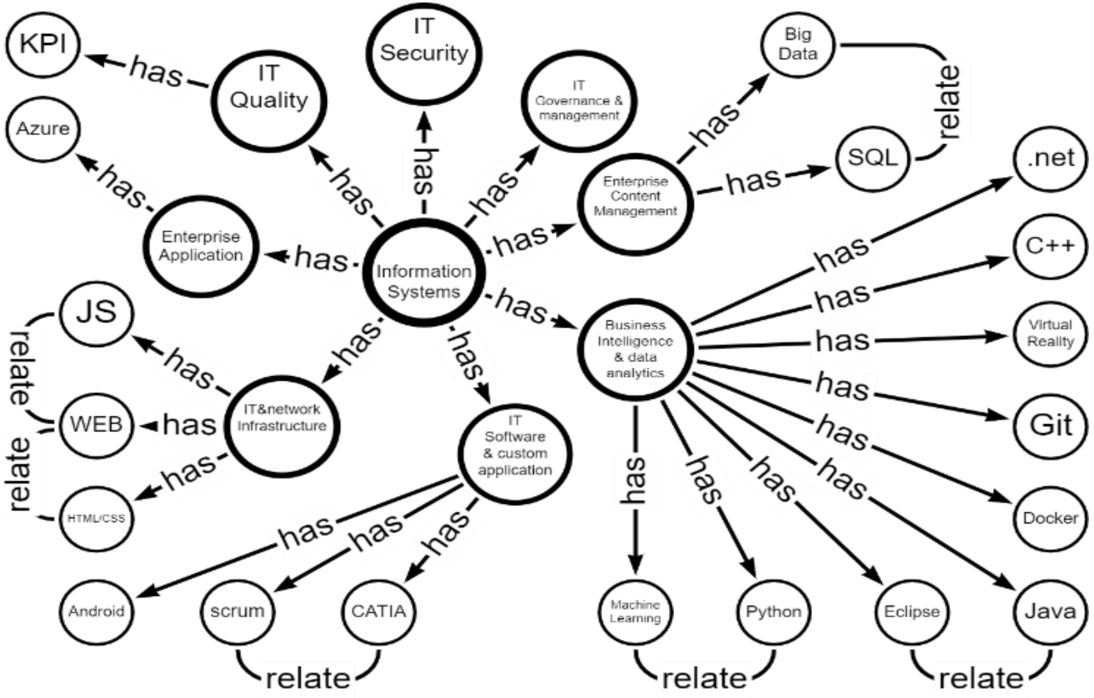
Proposed Knowledge Graph

Label - part of a group:
- *Competence Keywords*

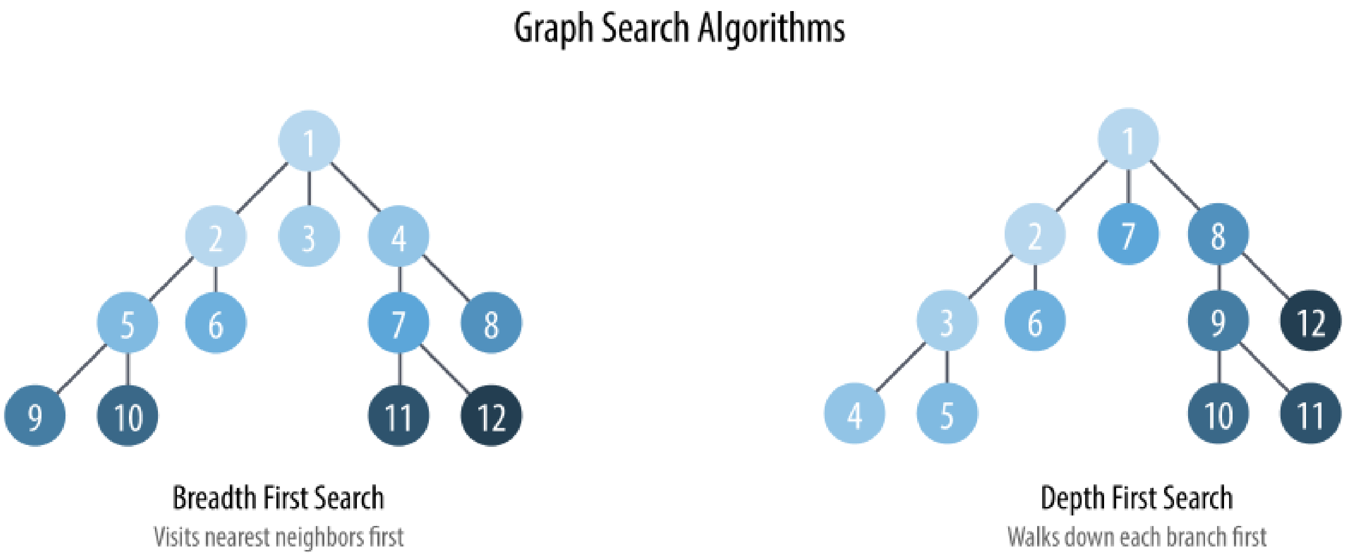
Relationship type:
- *Has*
- *Is a*

Scale-free network - a scale-free network is produced when there are power-law distributions and a hub-and-spoke architecture is preserved regardless of scale, such as in the World Wide Web.

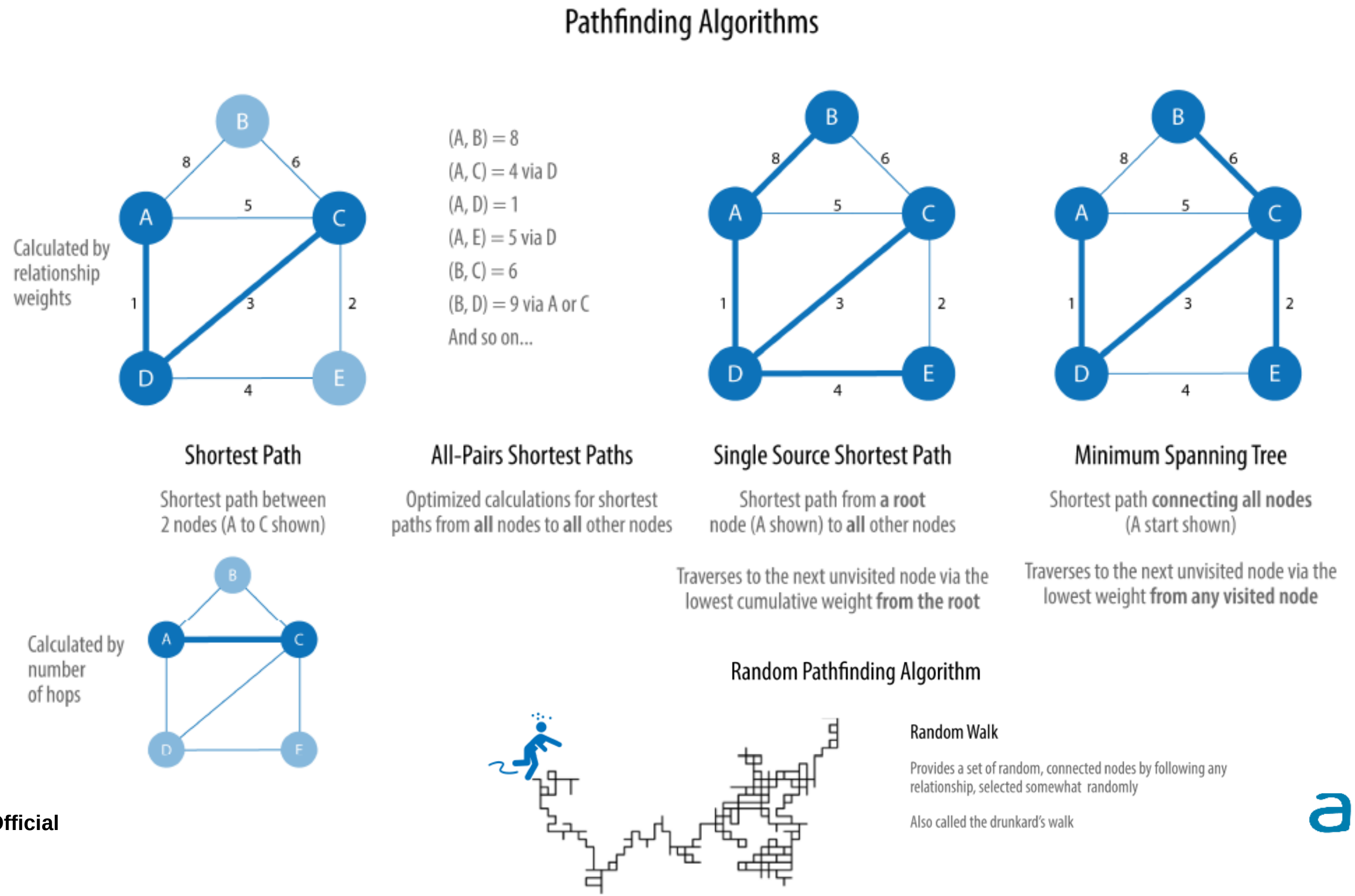
Flavors of Graphs:
- *Disconnected*
- *Unweighted*
- *Tree*
- *Sparse*
- *Bipartite – function and specialty*



Graph Search Algorithms



Pathfinding Algorithms



Overview of Pathfinding and Graph Search Algorithms

Algorithm type	What it does	Spark example	Neo4j example
Breadth First Search	Traverses a tree structure by fanning out to explore the nearest neighbors and then their sublevel neighbors	Yes	No
Depth First Search	Traverses a tree structure by exploring as far as possible down each branch before backtracking	No	No
Shortest Path	Calculates the shortest path between a pair of nodes	Yes	Yes
All Pairs Shortest Path	Calculates the shortest path between all pairs of nodes in the graph	Yes	Yes
Single Source Shortest Path	Calculates the shorest path between a single root node and all other nodes	Yes	Yes
Minimum Spanning Tree	Calculates the path in a connected tree structure with the smallest cost for visiting all nodes	No	Yes
Random Walk	Returns a list of nodes along a path of specified size by randomly choosing relationships to traverse	No	Yes

[Files · master · examples / Graph Algorithms · GitLab (oreilly.com)]

The State of the Art for KG Embedding Methods

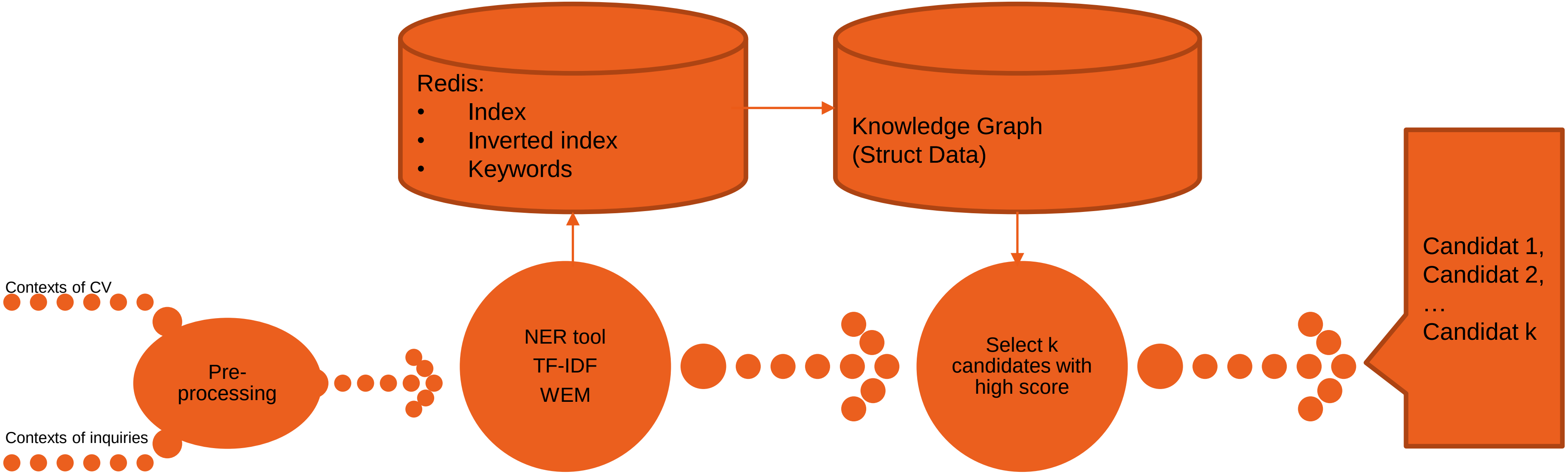
Entity embedding + Textual embedding + Visual embedding	Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 353– 362. ACM, 2016.
CNN + entity embedding + context embedding	Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. Dkn: Deep knowledge-aware network for news recommendation. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, pages 1835–1844. International World Wide Web Conferences Steering Committee, 2018
Three hop + entity embedding	Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pages 417–426. ACM, 2018
Meta-path + Bayesian	Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han. Personalized entity recommendation: A heterogeneous information network approach. In Proceedings of the 7th ACM international conference on Web search and data mining, pages 283–292. ACM, 2014
Meta-path + Bayesian + matrix factorization (MF) + factorization machine (FM)	Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lun Lee. Meta-graph based recommendation fusion over heterogeneous information networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 635–644. ACM, 2017
Meta-path + CNN + Co-Attention Model	Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1531–1540. ACM, 2018
Meta-path + RNN	Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. Recurrent knowledge graph embedding for effective recommendation. In Proceedings of the 12th ACM Conference on Recommender Systems, pages 297–305. ACM, 2018
Meta-path + LSTM	Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. Explainable reasoning over knowledge graphs for recommendation. arXiv preprint arXiv:1811.04540, 2018
Embedding + user preference = Joint Model	Cao Y, Wang X, He X, et al. Unifying Knowledge Graph Learning and Recommendation: Towards a Better Understanding of User Preferences[J]. arXiv preprint arXiv:1902.06236, 2019

4. Propositions

Search Engine (DBKDA paper 2021)
Recommender System

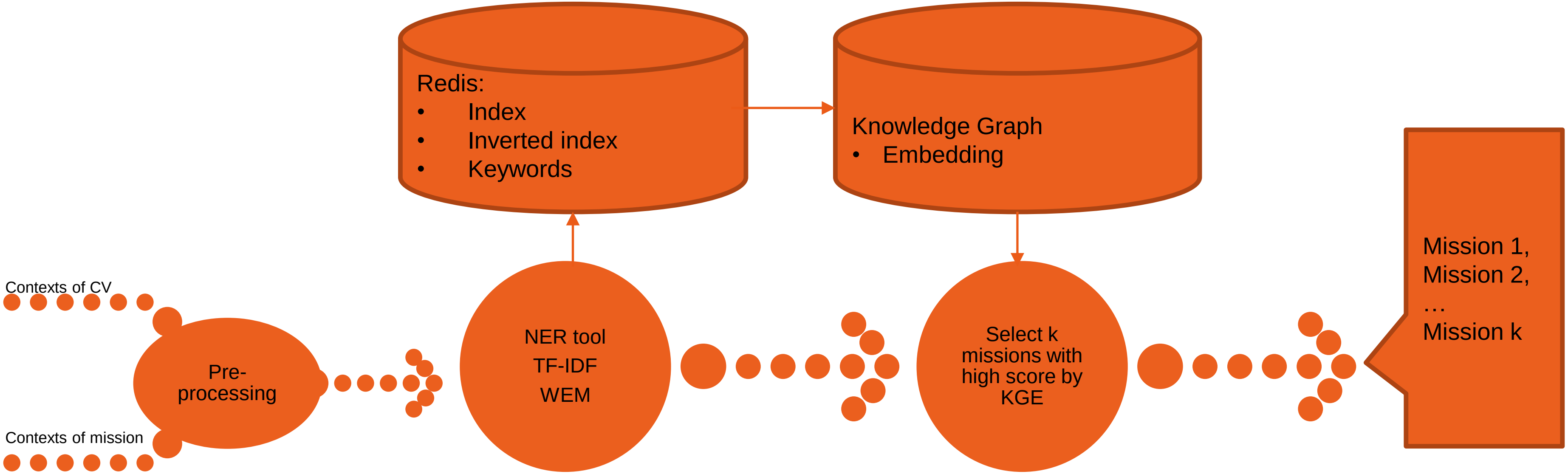
Proposed Approach of Search Engine

In the search engine, we take two kinds of files as input – CV and clients inquiries. We make a pre-processing to remove the ambiguous (for example, “JS” is “Java Script”). Then we use NER tool based on DistilBERT-base-multilingual-case to extract the competence keywords and TF-IDF to calculate the score of the competence keywords. We also use Weight Average Method (WAM) to calculate a global score. The scores are stored in Redis with index and inverted index. We select a list of candidates with high score as well as the related competence keywords from the Knowledge Graph.



Proposed Approach of Recommender System

In the recommender system, we take two kinds of files as input – CV and mission. We make a pre-processing to remove the ambiguous (for example, “JS” is “Java Script”). Then we use NER tool based on DistilBERT-base-multilingual-case to extract the competence keywords and TF-IDF to calculate the score of the competence keywords. We also use Weight Average Method (WAM) to calculate a global score. The scores are stored in Redis with index and inverted index. We need to propose an adapted KGE method in order to choose a list of missions with high score.



Perspectives of the Knowledge Graph

Dynamic Recommendation	It is natural to integrate other types of side information and build a KG for dynamic recommendation
Multi-task Learning	It would be interesting to exploit transferring knowledge from other KG-related tasks, such as entity classification and resolution, for better recommendation performance
Cross-Domain Recommendation	It could be promising to follow works by incorporating different types of user and item side information in the user-item interaction graph for better cross-domain recommendation performance
Knowledge Enhanced Language Representation	It is promising to apply the strategy of knowledge-enhanced text representation in the new recommendation task and other text-based recommendation tasks for better representation learning to achieve more accurate recommendation results
Knowledge Graph Embedding Method	Another research direction lies in comparing the advantages of different KGE methods under various conditions
User Side Information	Considering user side information in the KG could be another research direction

Thank you for your attention !

Yan WANG
Capgemini Engineering, Direction of Research and Innovation (DRI)
yan.wang2@altran.com

Part of Capgemini