Application of Information Retrieval for automation of Risk Engineering research within unstructured data

Amin Karbassi, Ph.D. | Senior Casualty Risk Consultant
Insurance Data Science Conference - London | June 2023

Motivation

- Risk Engineering's task: identify P&C risk trends to support the underwriting and pricing of risks

What technical solution can be implemented to automate the process of assessing information particularly in dealing with **unstructured data**?

Application of Information Retrieval for Risk Engineering research l Amin Karbassi

Example...

National Transportation Safety Board's website

Application of Information Retrieval for Risk Engineering research l Amin Karbassi   CONFIDENTIALITY LEVEL: Public

Text embedding

- Vectors manageable by computers, to compute how similar two pieces of text are to each other

- Number of values in a text embedding—known as its "**dimension**"—depends on the embedding technique
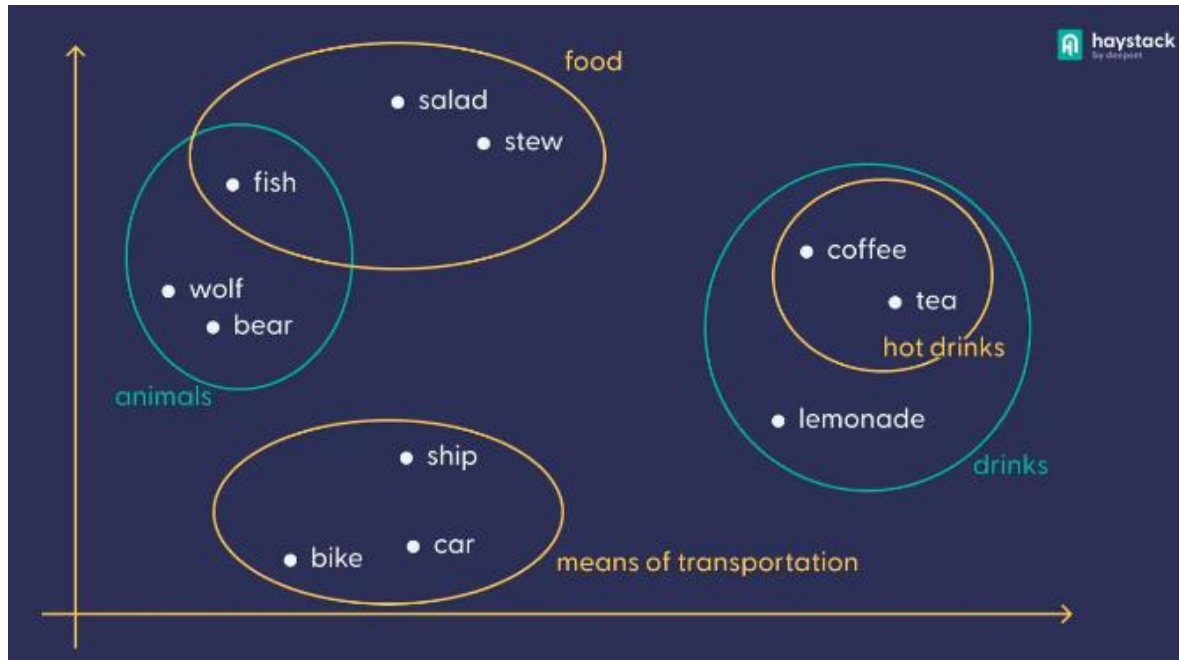

Photo source: haystack

Application of Information Retrieval for Risk Engineering research l Amin Karbassi   CONFIDENTIALITY LEVEL: Public

Vector Space Model
word2vec

- Created, patented, and published in 2013 by a team of researchers led by Tomas Mikolov at Google

- 'word2vec-google-news-300' model in the "Gensim" library

- Trained on the Google News dataset with about 100 billion words

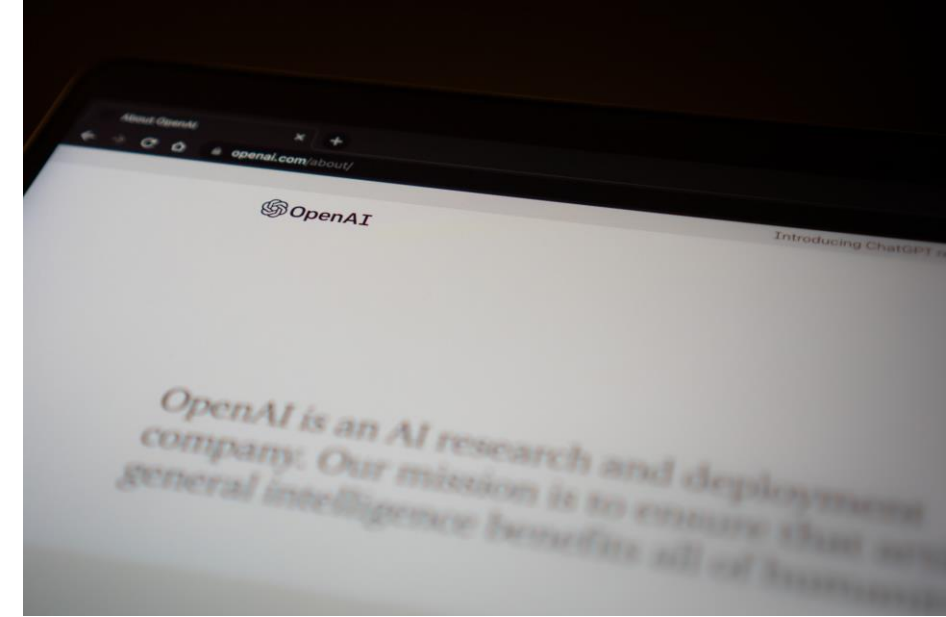|	Unnamed: 0	Location	Date	Title	What_Happened	Findings	Link
86	86	Hoxie, AR	Event Date\n8/17/2014	Railroad Accident Report: Collision of Two Uni...	?On August 17, 2014, at 2:28 a.m. central dayl...	NaN	https://www.ntsb.gov/investigations/Pages/DCA1...
122	122	Madison, IL	Event Date\n2/28/2012	Highway-rail grade crossing collision	On February 28, 2012, at 11:57 a.m., southboun...	We determined that the probable cause of the a...	https://www.ntsb.gov/investigations/Pages/DCA1...
101	101	Amarillo, TX	Event Date\n9/25/2013	Collision Involving Three BNSF Railway Freight...	?On September 25, 2013, at 4:17 a.m. central d...	NaN	https://www.ntsb.gov/investigations/Pages/DCA1...
62	62	Pine Bluff, AR	Event Date\n4/3/2015	Railroad Switching Services Employee Fatality	On April 3, 2015, at 9:23 p.m. central dayligh...	We determined that the probable cause of the a...	https://www.ntsb.gov/investigations/Pages/DCA1...
165	165	Elmwood Park, IL	Event Date\n11/23/2005	Commuter Train Highway - Railroad Grade Crossi...	?About 4:41 p.m. on November 23, 2005, the day...	NaN	https://www.ntsb.gov/investigations/Pages/HWY0...
...	...	...	...	...	...	...	...

OpenAI's text embedding
(Alternative model)

■ Embedding model ID: text-embedding-ada-002 (as of May 2023)
- replaces five separate models for text search, text similarity, and code search
- outperforms previous most capable model, Davinci, at most tasks, while being priced 99.8% lower

■ Highest accuracy across all machine learning models (tested on Amazon fine-food reviews dataset)
- Word2vec
- MPNet
- GloVe

Name	Dimension
gpt_3	1536
mpnet	768
word2vec	300
glove	300

■ Privacy issue for internal documents!

Application of Information Retrieval for Risk Engineering research l Amin Karbassi   CONFIDENTIALITY LEVEL: Public

Casualty Risk Consulting

Global Asset Protection Services, LLC, XL Catlin Services SE and their affiliates ("AXA XL Risk Consulting") provide loss prevention and risk assessment reports and other risk consulting services, as requested. In this respect, our property loss prevention publications, services, and surveys do not address life safety or third party liability issues. This document shall not be construed as indicating the existence or availability under any policy of coverage for any particular type of loss or damage. The provision of any service does not imply that every possible hazard has been identified at a facility or that no other hazards exist. AXA XL Risk Consulting does not assume, and shall have no liability for the control, correction, continuation or modification of any existing conditions or operations. We specifically disclaim any warranty or representation that compliance with any advice or recommendation in any document or other communication will make a facility or operation safe or healthful, or put it in compliance with any standard, code, law, rule or regulation. Save where expressly agreed in writing, AXA XL Risk Consulting and its related and affiliated companies disclaim all liability for loss or damage suffered by any party arising out of or in connection with our services, including indirect or consequential loss or damage, howsoever arising. Any party who chooses to rely in any way on the contents of this document does so at their own risk.

AXA, the AXA and XL logos are trademarks of AXA SA or its affiliates. © 2023.

Know You Can