Unveil the obscure network of company and location data using smart algorithms

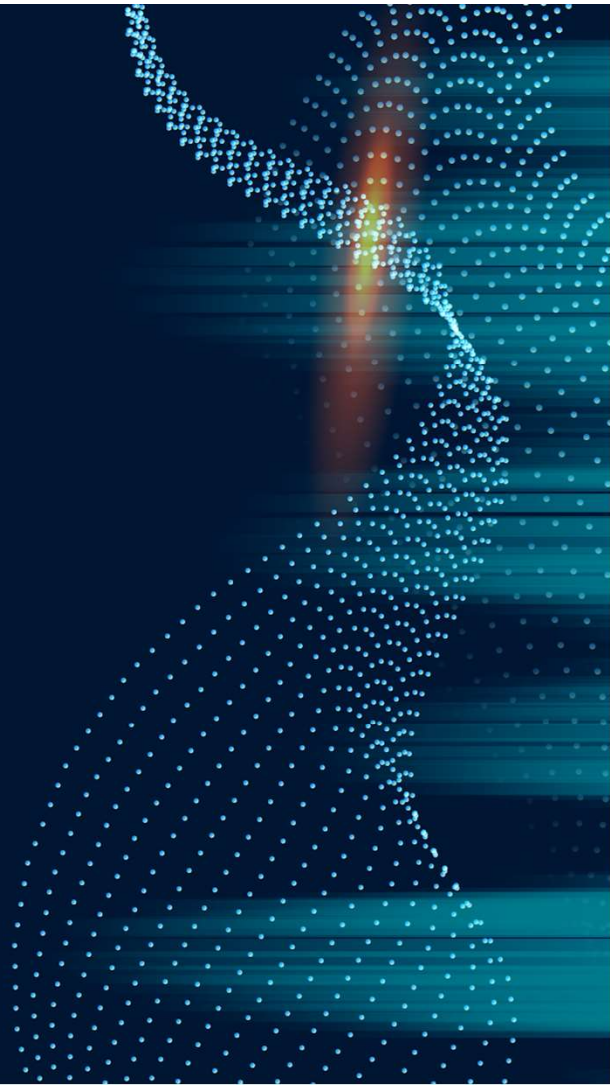
Team 2: Susann, Ojasvi Prakhar, Marco, Stephanie

TABLE OF CONTENTS

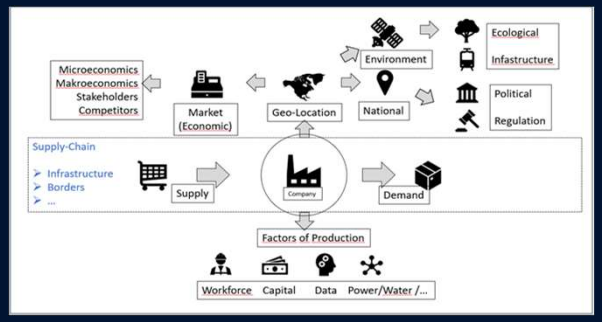
01 OVERVIEW

PROBLEM CHARACTERISATION SPRINT 1

1. Defining features in which a company and its locations (e.g. factories, warehouses, sales points) can be interlinked.
2. Identifying free or paid source of information, where data supports applying an ontology to the actual companies and buildings.
3. Select methodologies for the implementation



Model & Clustering according to factors & Data sources



Model is self made based on PESTEL, Porter 5, Fourastie and St. Gallen Mgmt Modell

Cluster	Need	Explanation	Data Source
Factor of Production	Work force	Educational level Age of population	CIA factbook
	Power supply Water access	Reliability of Power brid Access to water	Global Energy Network Institute
Supply Chain	Access to transportation	Roads, Railway, Seaport, Airports	Google Maps, Satellite images, Data of shipping companies
Political / Regulation	Political set up and local regulations	Political stability, government form, corruption	CIA factbook, Transparency International Index
Economic	Economic Growth	GDP	CIA factbook
	Rates	Exchange rate, Interest rate Inflation rate, Unemployment rate, average salaries, industrial automation	Dun Bradstreet, CIA factbook, Worldbank, statista
	Tax	Corporation tax, Capital gain tax	WTO

Focus on Location data

SELECTION OF COMPANY / COUNTRY

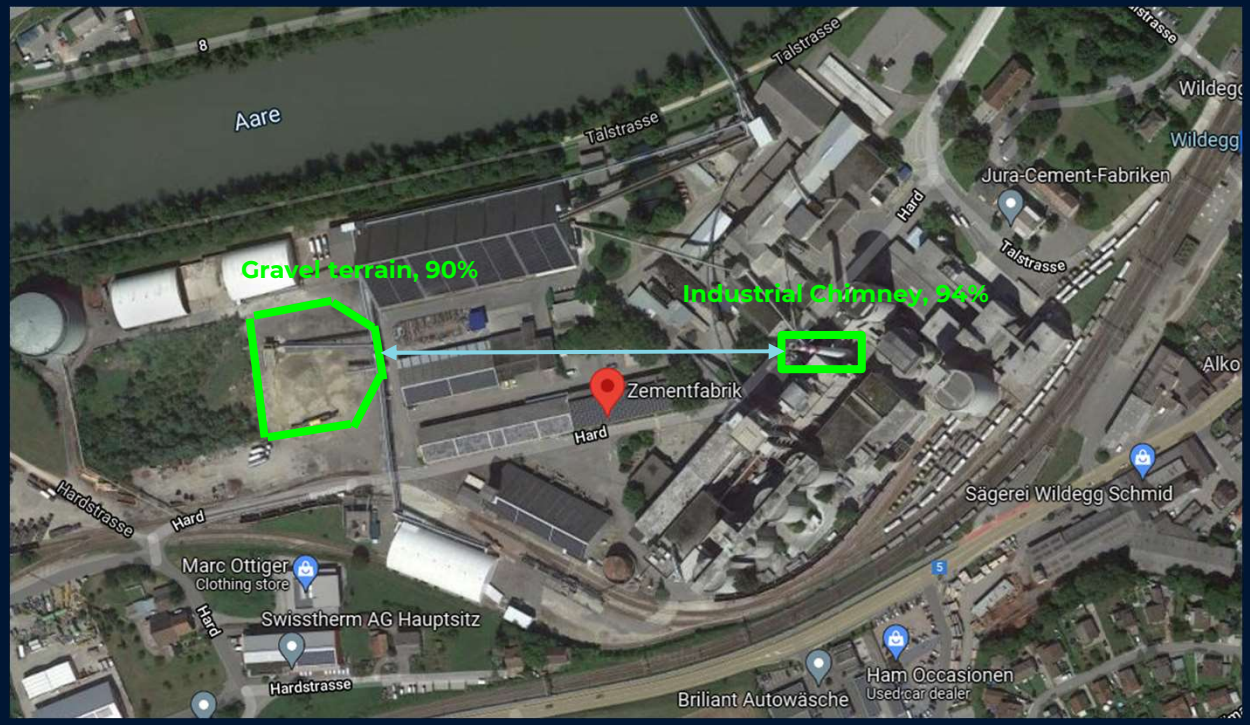
- At least 2 way footprints inside model fourastie
- Worldwide availability and several locations
- Listed at stock exchange to have accessible data
- Company ensures challenges with the name (e.g. through M&A transactions)
- Industry selected ensures images for satellites can be differentiated
- Focus on sustainability (“Green cement industry) and have Importance for the infrastructure of a country
- Biggest supplier worldwide
- Cement industry: 1st producer is China, 2nd India
- Exclusion of China for sample set due to availability of data

Sample: LaFarge Holcim

- M&A 2014 Swiss company, brand in India Ambuja Cement Limited
- Primary Sector: Mining of gravel and cement
- Secondary Sector: Production of concrete
- Countries: Switzerland, Colombia, India

CEMENT INDUSTRY: Bird’s Eye View

Gravel site and industrial chimney, each with high confidence detected on satellite photo and terrain views. In combination and close to an urban or industrial area, railway and / or main street access, water source nearby, as well as several warehouses, silo and outdoor conveyor belts, very likely pointing to a cement factory.



SCALED & ACCESSIBLE DATA SOURCES

- Company information: Dun & Bradstreet, Importyeti.com, Company websites, Chambers of commerce

  (Unique DUNS number in paid version of DnB allows overcoming same naming)

- Blacklisted companies: Trustpilot.com, ... Question to Team 2: Your 1st Sprint II goal, is this to focus on "reading from the signs" for example or also figure out based on the other satellite imagery you outlined before?

- Question 2 to Team 2: What is the viability of the solution parts focusing on imagery? That can get "expensive" in terms of processing lots of data.

-
- Relationship and hierarchies of companies: DnB

02 METHODS

METHODS

Searchable Ontology

An ontology built with a scraper that Swiss Re can use for searching for a particular company within the database.

Analysis from top view

Applying cutting-edge object detection methods to satellite, drone, airplane map APIs

Analysis ground view

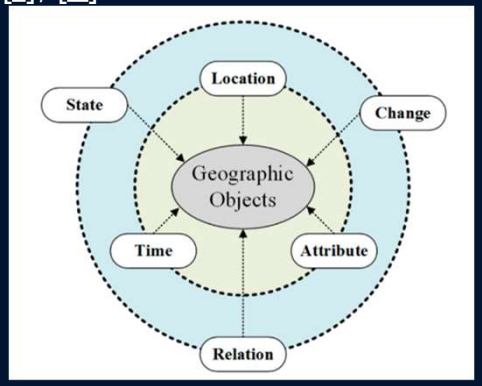
Combining Natural Language Processing (NLP) text extraction with Computer Vision (CV) methods to Streetview or Geological data

Visual representation of Ontology via Geographical Knowledge Graph (GeoKG)

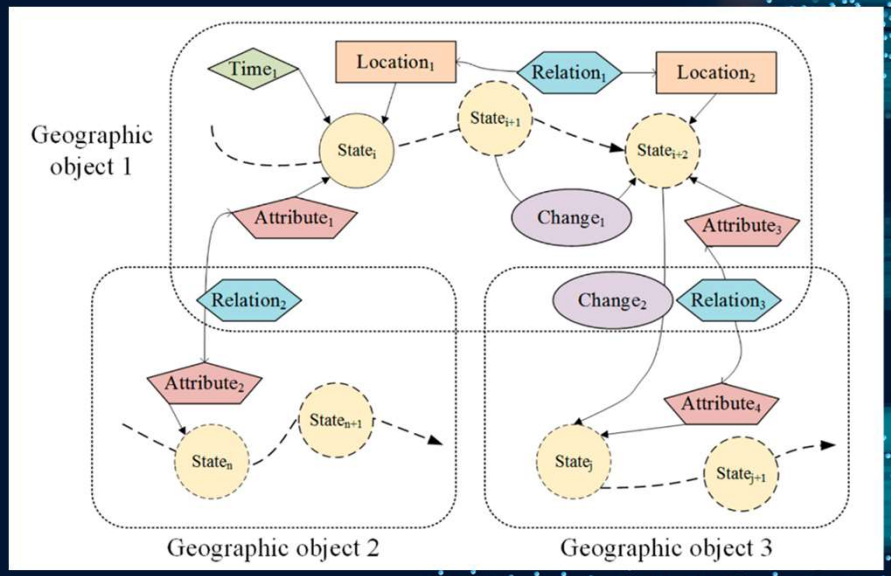
Using the concept of Spatial relations and sameness among different geographical objects, we build a model that studies the relationship between the existing and possible components of the landscape as well as the situation.

[1], [2]

(a)



(b)



a) Six elements that represent a GeoKG Object
b) The interplay between multiple GeoKG objects

Searchable Ontology: Analysis

The business address information can be extracted from the Dun&Bradstreet with manufacturers, subsidiaries etc., or also from ImportYeti website with a web scraper or can merge it with a Selenium script. Then can plot these locations on a map or a heat-map to visualise the locations or the supply.

Strength

- Easily Searchable and less coding required
- NLP methods allow for hierarchy and relationships retrieval
- Can be hosted on the cloud with differential access
- Faster searching because the data is already loaded
- Highly scaled, works globally

Weakness

- Maintaining a database can be a bit less user-friendly
- Companies with the same name or lots of name changes
- Can take up a lot of space to store over a period of time

Analysis from Top View

With object identification algorithms, such as recent YOLT or YOLO versions, CNN and transformer-based solutions applied to satellite images or drone photographs in various spectra we can also provide locations of factories or tech parks that have a fence or wall around, and through this multiple buildings or offices of a company inside it. Similarly, for other domains, such as mining, power networks or factories using transfer learning applied to foundational models, or self-learning.

Strength

- Allows finding buildings and objects not registered or covered in any database
- Doesn't take up a lot of space since it doesn't store data, unless specifically desired
- Updated daily, works globally
- Provides further visual and more spectral data for other risk analyses too

Weakness

- Takes more time to build and more code or scripts
- Can be slower because API or inference may run each time
- Limited to buildings and objects or industries that can be uniquely identified or attributed

Analysis Ground View

Can perform CV on Google Earth/Street View together with NLP multi modally via models we have built using for instance the Sagemaker APIs to extract company names from the building names or entrance inscriptions to match and complete the list of buildings that may be away from others or detached.

Strength

- Our modern computer vision text extraction methods are very precise
- More extensive and accurate than from satellite view
- Very detailed, allowing for further risk factor analyses
- Works for almost all industrial areas

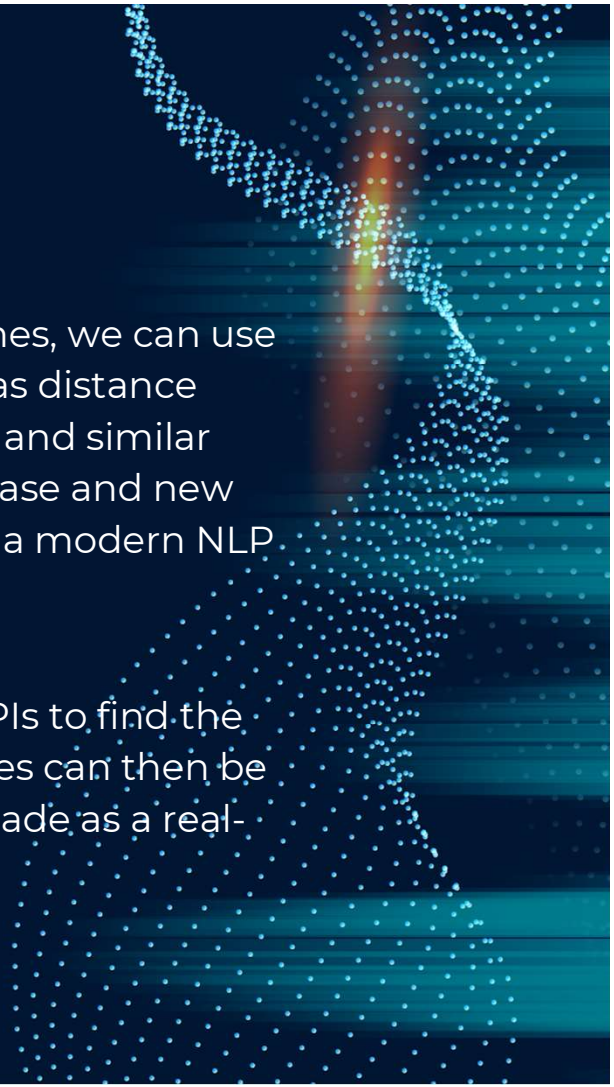
Weakness

- Limited to where street image data or other image data is available
- Less frequently updated
- Not showing company relationships

Address Text Normalisation
(Secondary Problem)

- For the secondary problem of normalising address names, we can use fuzzy logic for address matching with the Levenshtein as distance metric for a probability score for how close they match, and similar address matching methods, such as the SwissRe database and new web APIs. The accuracy can also be enhanced further via modern NLP or bi-modal pattern matching methods.

- This can be combined with the Google Maps or Bing APIs to find the corresponding location and coordinates. The coordinates can then be stored in a database or ontology. This process can be made as a real-time service using Python scripts.



03 OUTLOOK

Our Solution

NLP Ontology

- Very complete, global
- Fast, readily available
- Not expensive
- Automatic update alerts
- Fraud prevention
- Better risk assessment

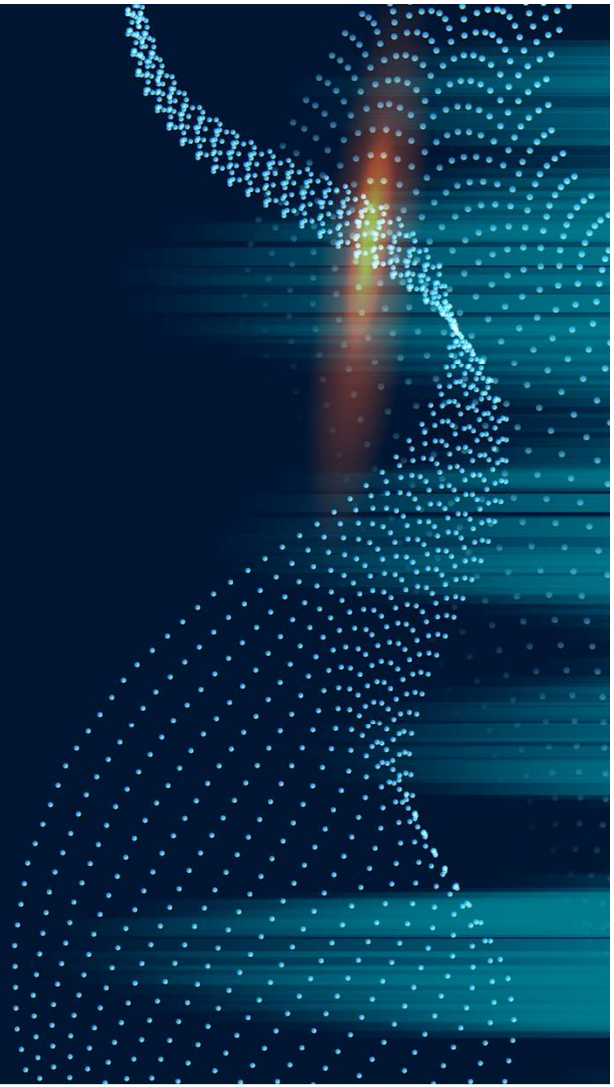
Computer Vision Extension

- Can complement the missing entries globally
- Images updated daily
- Independent analysis for further fraud prevention
- Unmatched risk assessment

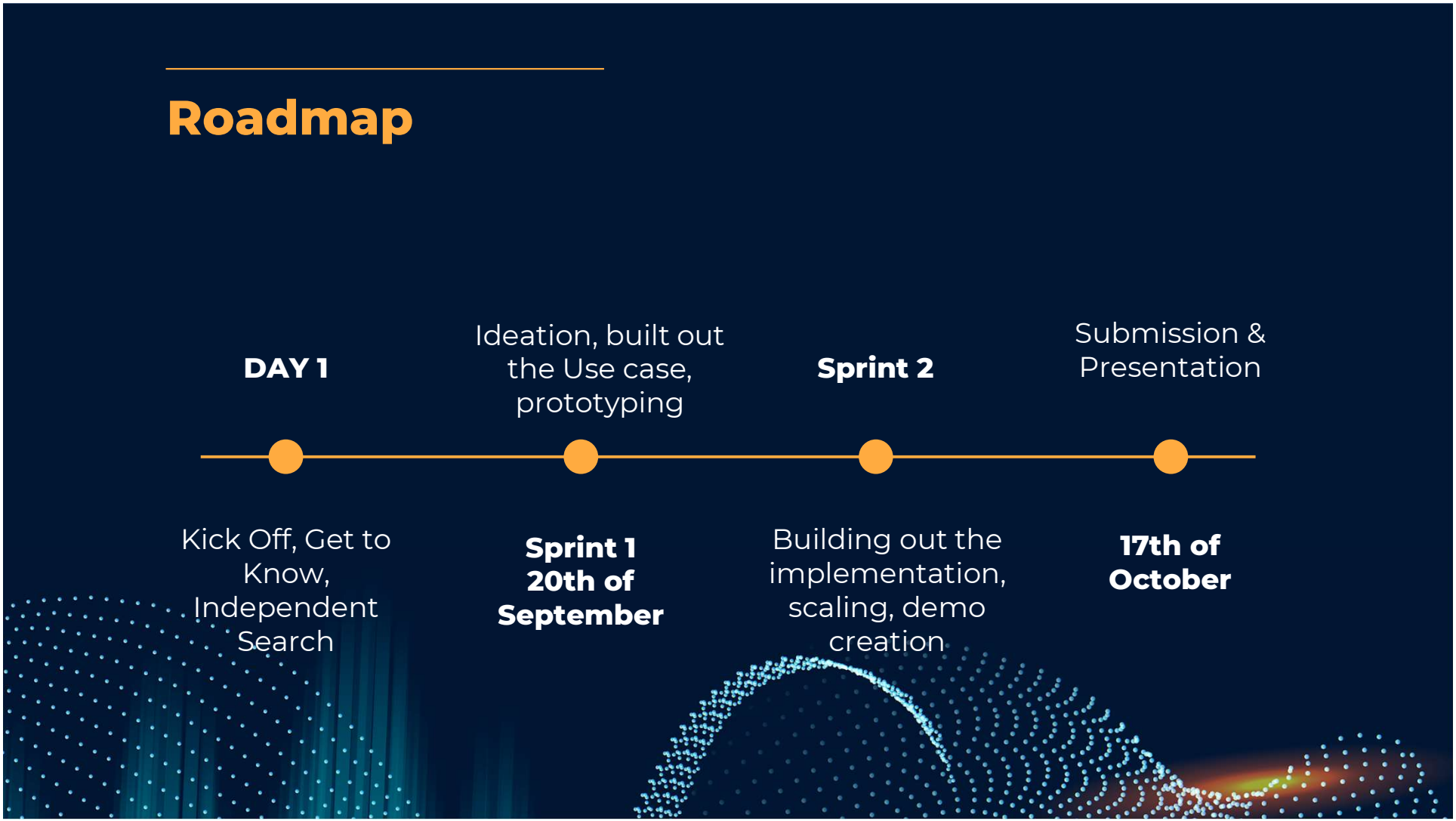
Outlook for SPRINT 2

Our goals for sprint 2 are:

- Implement the Vision API to extract company names from Google Street View and Google Earth and other sources
- Build a searchable ontology for Swiss Re to abstract the API calls and matching processes in the background
- Perform text normalization exercise to match the company address to a single address irrespective of spelling.



Roadmap


DAY 1
Kick Off, Get to Know, Independent Search
Sprint 1 20th of September
Ideation, built out the Use case, prototyping
Sprint 2
Building out the implementation, scaling, demo creation
17th of October
Submission & Presentation

THANKS!

Do you have any questions?

Team 2

Question to Team 2: Your 1st Sprint II goal, is this to focus on "reading from the signs" for example or also figure out based on the other satellite imagery you outlined before?
Question 2 to Team 2: What is the viability of the solution parts focusing on imagery? That can get "expensive" in terms of processing lots of data.


BackUp

Overview Cement Industry

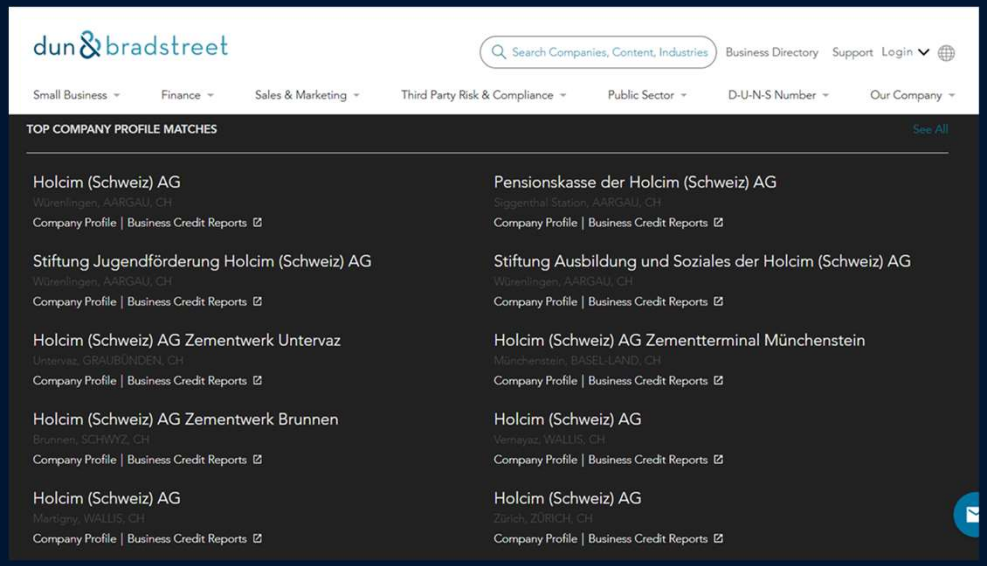
Top 10 Largest Cement Companies in the World 2020

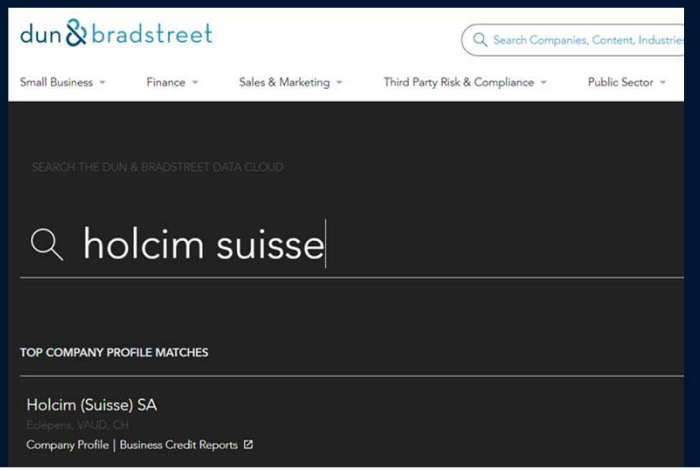
Who are the top 10 cement companies in the world in 2020? The following is a list of the world's largest cement companies ranked by cement production in million tonnes.

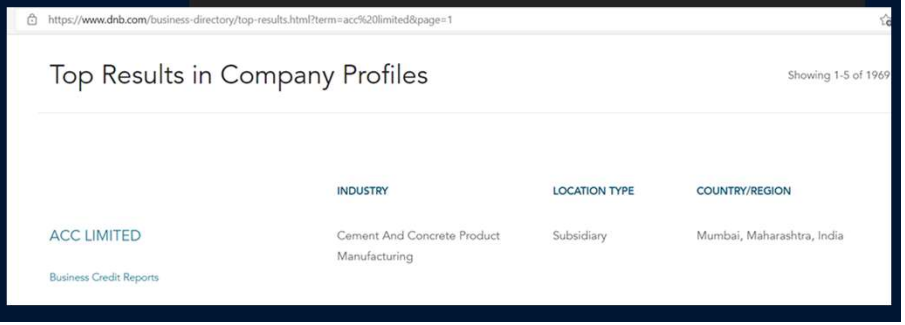
Rank	Company	Annual Cement Production (Mt/yr)	Cement Production Capacity (Mt/yr)	Country
1	LafargeHolcim	286.6	386.5	Switzerland
2	Anhui Conch Cement	217.2	288.0	China
3	CNBM	176.22	406.0	China
4	Heidelberg Cement	121.11	129.0	Germany
5	Cemex	87.09	93	Mexico
6	Italcementi	76.62	77.0	Italy
7	China Resources Cement	71.02	78.3	China
8	Taiwan Cement	63.72	69.0	Taiwan
9	Eurocement	45.18	50	Russia

Ideas of companies LafargeHolcim

- Test Duns & Bradstreet website - three cement manufacturer in Switzerland and ACC limited in India available
- Google Maps: Search with name Holcim Schweiz AG only 2 hints







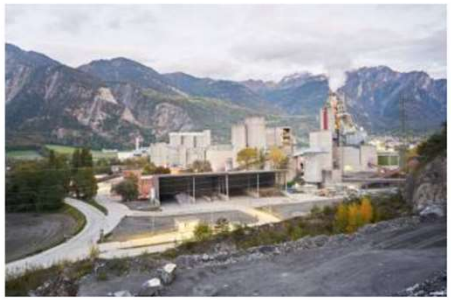
Lafarge Holcim

Unsere Zementwerke

Holcim (Schweiz) AG ist mit jährlich gut 2,5 Millionen Tonnen Zement aus den drei Werken Eclépens, Siggenthal und Untervaz die grösste inländische Zementproduzentin.

Zementwerk Siggenthal

› Mehr erfahren



Zementwerk Untervaz

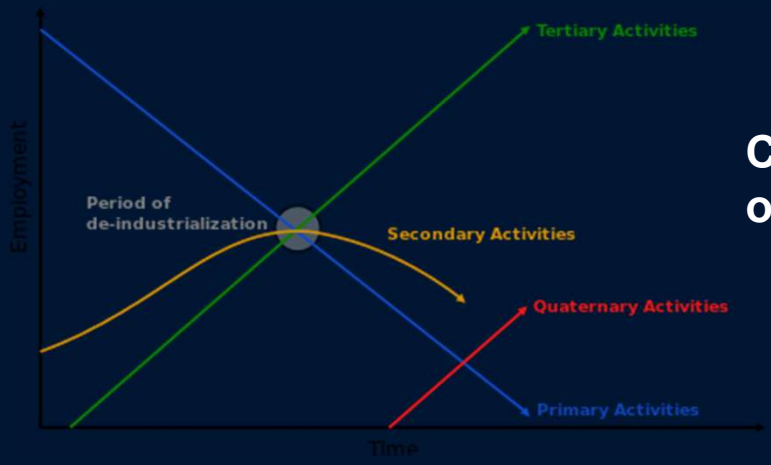
› Erfahren Sie mehr

Zementwerk Eclépens

› Erfahren Sie mehr (in Französisch)

THREE SECTOR MODEL FOURASTIE TO WEIGHT RELEVANT FACTORS OF PESTL

- Primary: involves the retrieval and production of raw materials, such as corn, coal, wood or iron.
- Secondary: involves transformation of raw or intermediate materials into goods, as in steel into cars, or textile into clothes.
- Tertiary: involves the supplying of services to consumers and businesses, such as babysitting, cinemas or banking.
- “Quaternary" and “quinary” service sectors: economic activity in the hypothetical quaternary sector comprises knowledge- and info-based services, while quinary services include industries related to human services and hospitality.

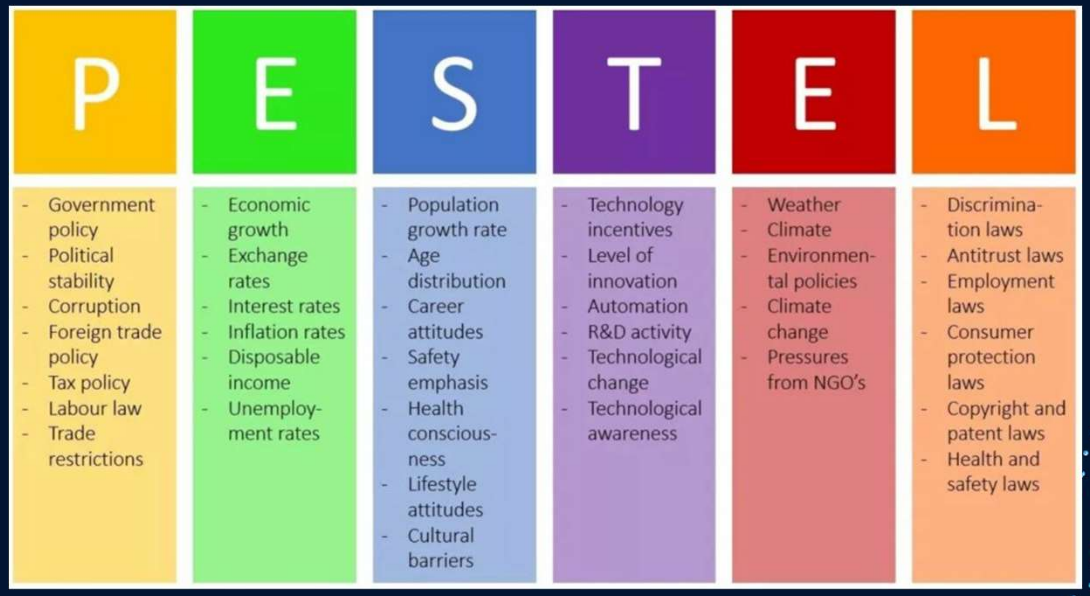


Company of interest has impact on Primary & Secondary Sector

Economic sector - Wikiwand

Influencing factor for a company to evaluate risks

- PESTEL- Political, Economic, Socio-cultural, Technological, Environmental geographic and Legal.



PESTEL Analysis (PEST Analysis) EXPLAINED with EXAMPLES | B2U (business-to-you.com)

Data sources for referring links

Database / Source	Type of Information
Google maps or Google Street View, google earth	• Computer vision on satellite imagery or Street view extraction
maps.me, yelp, bing	• Computer vision on satellite imagery or Street view extraction
Duns and Beadstreet dnb.com	• comprehensive business data and analytical insights to power today's most crucial business needs.
Companies website holcim.com	• Subsidiary and parent company information, address code • annual/ quarterly reports
chamber of commerce moneyhouse zefix.com	• information official listed companies example Switzerland
placekey.io/blog/adress-matching	• Ideas on way how to solve it
importyeti.com	• supplier and companies/manufacturer information
The World Factbook Archives - The World Factbook (cia.gov)	• Information on countries Switzerland - The World Factbook (cia.gov) • Colombia - The World Factbook (cia.gov), India - The World Factbook (cia.gov)
Indian Cement Industry Analysis | IBEF	• Indian Cement Industry Analysis | IBEF
Top 10 Cement Companies in the World 2020 | Global Cement Industry (bizvibe.com)	• Top 10 Cement Industry worldwide
Global Energy Network Institute	• National Energy Grid - Index - Global Energy Network Institute - GENI is the highest priority objective of the World Game (R. Buckminster Fuller).

OVERVIEW

PROBLEM CHARACTERISATION

Linking companies to location data

COUNTRY & INDUSTRY

Switzerland, India, Colombia
&
Cement Industry:
LaFarge Holcim