Stefan Morgenweck
Benjamin Durupt

I/O.VATE

A comprehensive analysis of the innovative output in the Hong Kong Special Administrative Region from 2006 to 2018

Contents

Motivated to Shed Light Into a Dense Cave

01. **Description: what?**

- ▶ Patents protect inventors and enable economies to flourish
- ▶ The principles of protection and disclosure result in an openly accessible body of knowledge

02. **Rationale: why?**

- ▶ There is a vast amount of patents. Existing classifications are not enough to look to produce insights

03. **Target Audiences: who?**

- ▶ Capital givers
- ▶ R&D Teams
- ▶ Individuals
- ▶ Bodies of education

04. **Commercialization**

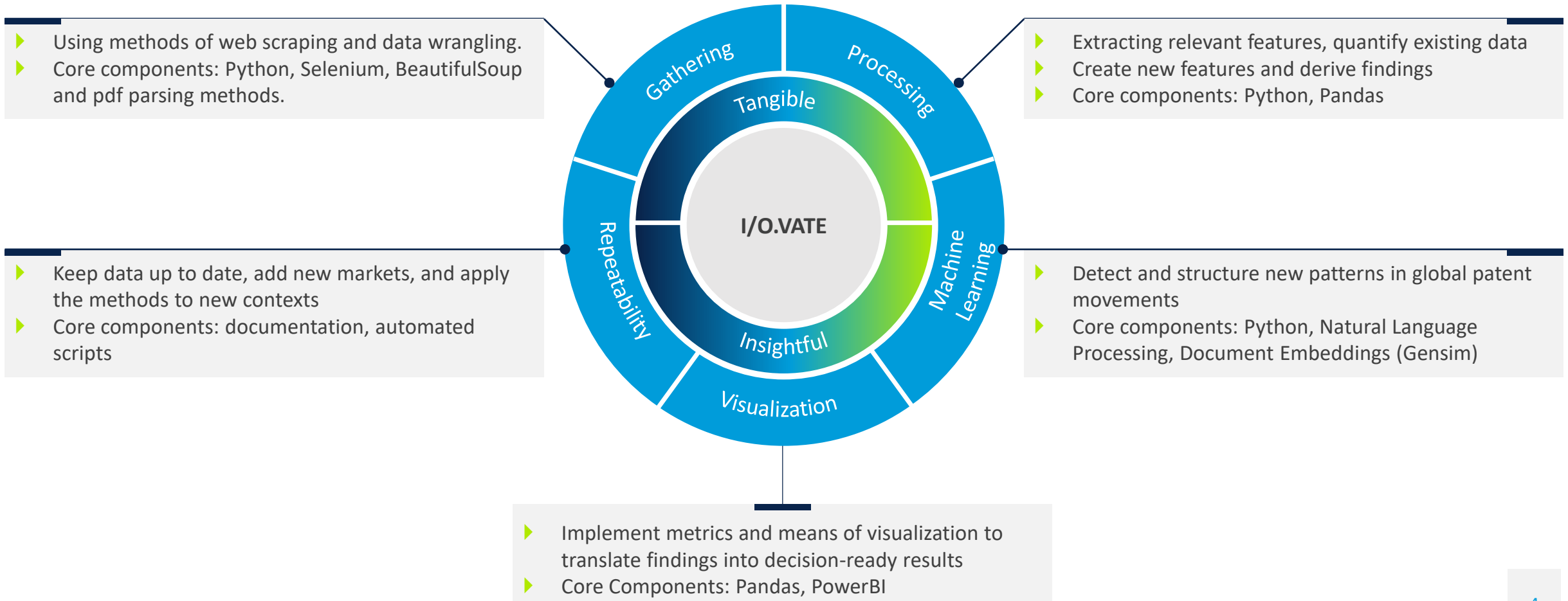- ▶ The refined process bears business potential

The Question?

"Are the patents that are being filed in the Special Administrative Region of Hong Kong (HKSAR) following underlying patterns that go beyond a usual classification of industries and such features on the surface?"

"How can a niche market with high potential be described, based on this assumption? "

Holistic Approach to Get Tangible, Insightful Results

By combining a variety of analytical approaches, we aspire to untangle the ball of wool and locate some Golden Fleece.

- Using methods of web scraping and data wrangling.
- Core components: Python, Selenium, BeautifulSoup and pdf parsing methods.

- Extracting relevant features, quantify existing data
- Create new features and derive findings
- Core components: Python, Pandas

- Keep data up to date, add new markets, and apply the methods to new contexts
- Core components: documentation, automated scripts

- Detect and structure new patterns in global patent movements
- Core components: Python, Natural Language Processing, Document Embeddings (Gensim)

- Implement metrics and means of visualization to translate findings into decision-ready results
- Core Components: Pandas, PowerBI

Gathering

Processing

Tangible

I/O.VATE

Repeatability

Machine Learning

Insightful

Visualization

Today vs. Tomorrow
Goals, risks, rewards and aspirations

Today

- The global patent databases are deeply nested
- Information retrieval is difficult and non-exhaustive
- Mere category classifications don't help to source valuable information
- There are underlying patterns that are not visible at plain sight

Tomorrow

- The massive corpus can be used to identify key players
- Changes over time show tendencies in all granularities: economies, industries, companies, even single individuals
- The efforts of research and development can be adjusted to what is happening left and right
- Matching capital supply to demands based on tangible indicators

Current Status & Next Steps

Finding the data sources took one week.

Currently we have applied our model and invoked vectors. Next, we will try to add new documents and see how well it performs.

Finding our data

Applying model

| 01. | 02. | 03. | 04. |

Preparation

Make it tangible

As we have understood, this may comprise 80% of our time, and it did.

Visualization and documentation of our findings is the final step.