ontotext

In the era of bots serving as investment vehicles and plenty of platforms offering tools for automated investing, could it be that the strength of investment strategies soon becomes a matter of computing power and algorithms well-built?

With more than 8 million companies and 1.5 million financial transactions in mind, sorry – in a graph, it might well be.

Approaching investment opportunities with 8 mln companies in mind (or in a graph)

The mid 19th century, writes Steven Marks in his book *The Information Nexus*, saw the coming of new information technologies on Wall Street that saved time and lowered transaction costs. The messenger who back then gathered prices by running back and forth between brokerage houses and the stock exchange lost out to the stock ticker – a device that printed the fluctuating prices as they streamed continuously over the telegraph lines. [cit. p. 167, The Information Nexus]

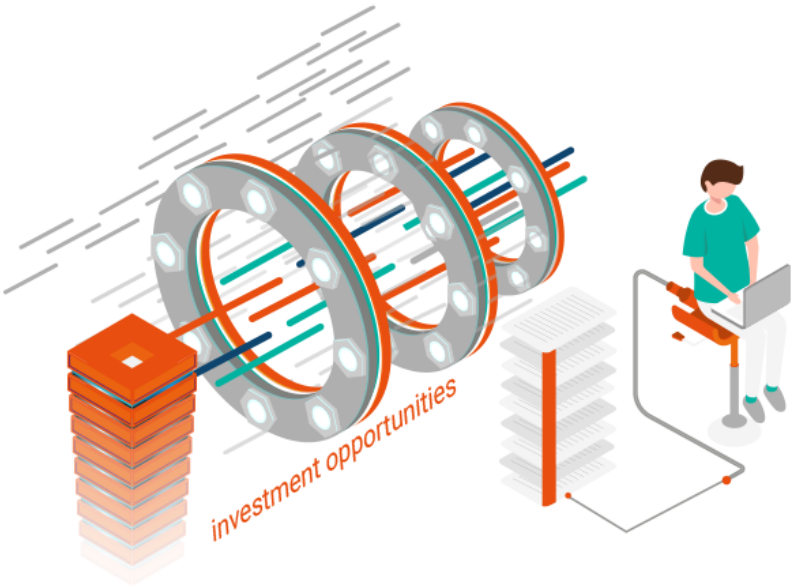
Today the advancement of the information nexus continues. New technologies and new means of processing and analyzing data accelerate the circulation of information about capital markets and business opportunities and reconfigure the way markets work. And we might speculate that these techniques will take over the investor's task of analyzing and spotting opportunities.

But we will be wrong. They will only shave off X-hours of time for processing information.

It is true that having automated investments is a lucrative option. However the complexity of the problem is too high to allow complete automation. As with the entire man-machine-symbiosis discourse, the matter of outsourcing typically human activities (in this case considering investment options and finding the best among them) to computer programs seems to be best approached with reframing the question of automation: *The key question shouldn't be "Whether and how to automate it all?", but rather "What part of it can be automated and what techniques will serve best the automation?"*

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext



What part of outsourcing human activities can be automated and what techniques will serve best the automation?

In the case of an algorithm for finding investment opportunities, the answer is straightforward: The part that will be automated is the part of narrowing down interesting investment opportunities from millions of companies worldwide. And the techniques to be used are data mining techniques to identify patterns of investment opportunities within a custom dataset containing around 8 million companies and investors and 1.5 million investment events.

Long story short, it takes a well-built knowledge graph and a set of algorithms to mine it.

In more detail, reducing the work of sifting through millions of companies and returning a list of interesting investment alternatives for an investor to consider is an active interplay of linking data, algorithmic operations and human expert's analysis.

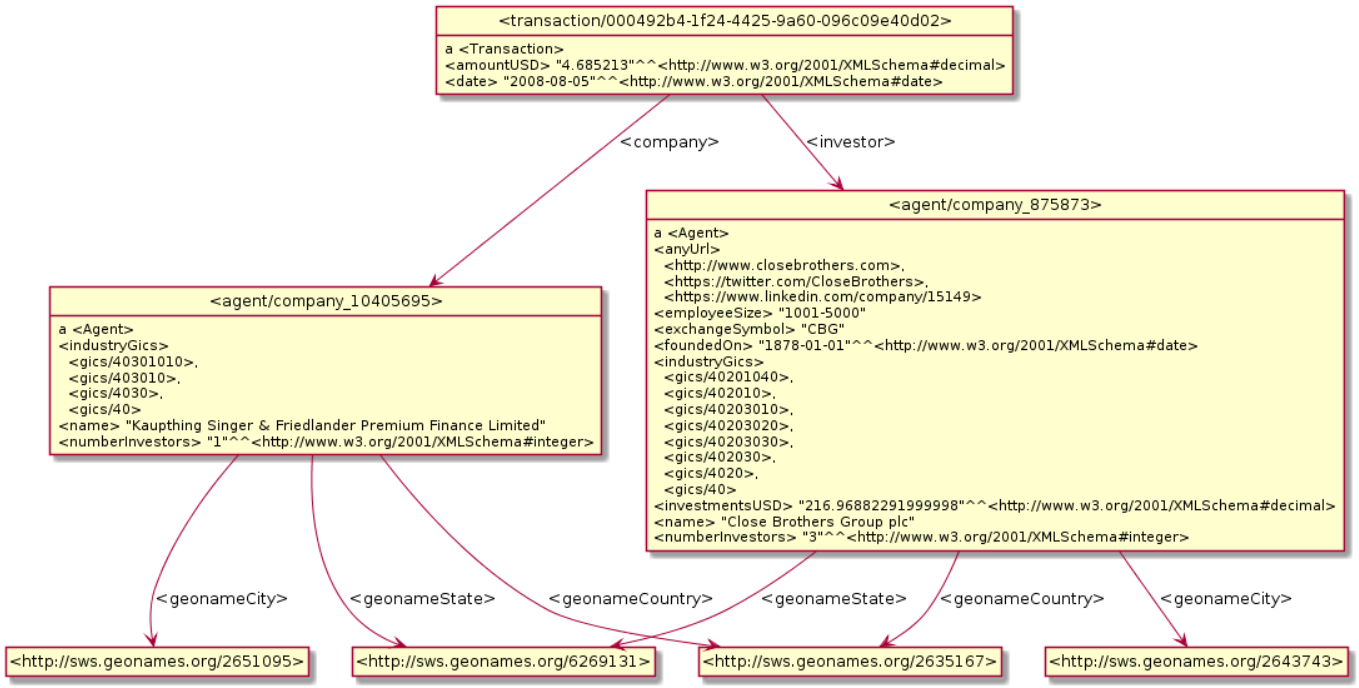
Below is the story of how Ontotext put into practice the idea to help investors successfully find investment options without having to sift through millions of options. You can see the full scientific article "Company Investment Recommendation Based on Data Mining Techniques" published in Springer (or in Google books) to learn how inductive logic programming, clustering and other techniques were applied and evaluated.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

In order to compute good investment opportunities and make it easier for investors to spot hits, Ontotext set to the task of identifying potential investment alternatives. The identification process was based on the historical data of investment events contained in a knowledge graph with over 8 million companies and 1.5 million financial transactions they built.

This knowledge graph was built by fusing a total of 5 datasets, containing features like the name of the company, rank, country, region, industry, etc. To build and continuously update such a big graph, without degradation of data quality, Ontotext used its GraphDB semantic graph database engine and an efficient semantic data integration pipeline, which allows regeneration, reconciliation, validation, refinement and indexing over night. One of the pillars of the calculations was the relationship between an investor and a company – the so-called investment event.

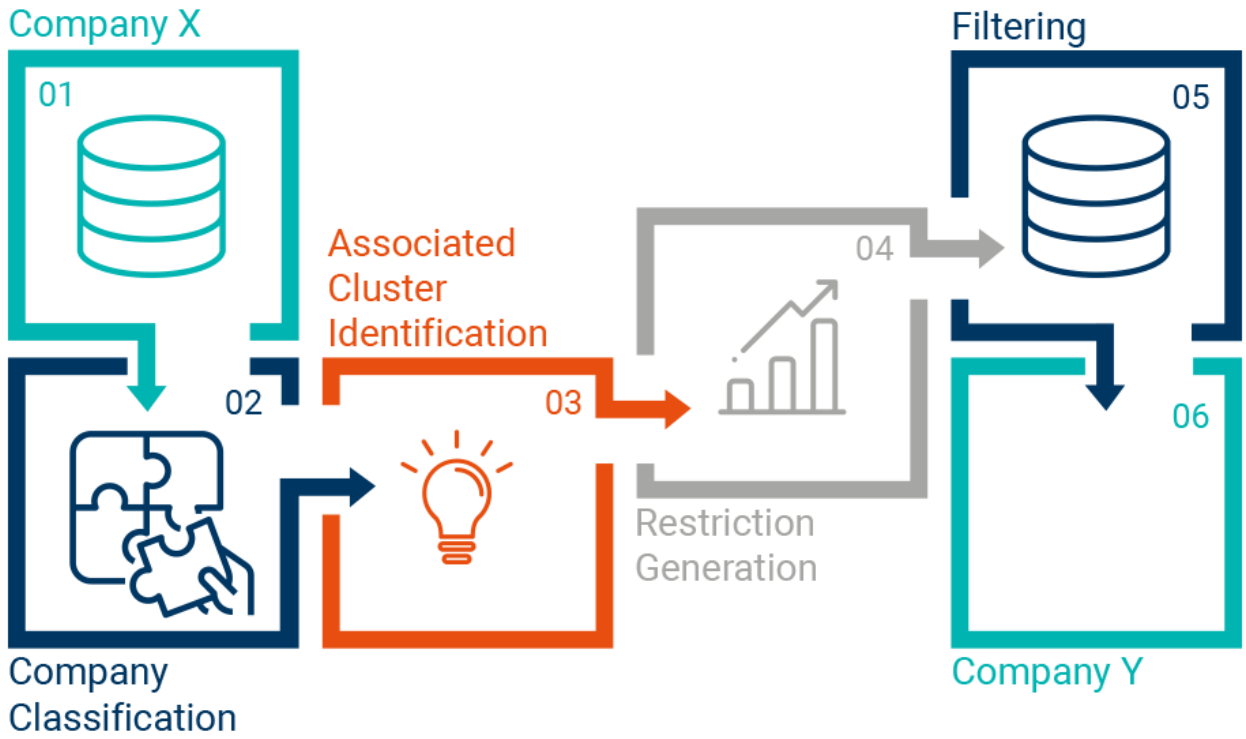


This very connection between company and investor through an investment event was what served as the basis for training machine learning algorithms that could further find and recommend data answering the specific criteria it had been given. In essence, the algorithms examined different portfolios of investments and look for certain patterns within them to

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

conclude what makes a certain company a good fit for a given investment portfolio.

INFOBOX: It takes 6 steps to mine an investment opportunity.



As regards to the methods, there were several crucial details in finding the most promising combination of algorithms that would serve investors in the pre-selection step of finding investment opportunities: the idea of non-personalization and the idea of indirectly associated companies.

When personalization is not an edge

With a wealth of information oxygenating market activities, the number of approaches is rising and the majority of the proposed solutions are personalized. This particular data mining approach however took the road less travelled – the non-personalized approach.

Knowing how important it is not to leave any trace of choices for investing, as this would be downright brainpicking, Ontotext aimed to develop a method that was non-personalized,

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

data-driven and unsupervised. Another reason for the lack of personalization was the fact that a personalized method would have made the calculations too complex. Here, the aim was to just model the behaviour of what any investor (or analyst) would do: search for companies by given criteria, trying to find his or her way towards the best investment opportunities.

When the association is indirect

One of the distinctive features in this automation of investment opportunities searches is the idea of investigating direct and indirect associations of company investments.

Direct association, also called frequent patterns, represents different sets of companies that appear together in the investment portfolios of multiple companies. Indirectly associated companies are again connected, but in the way Nike and Reebok are. Unlike typical association rules, indirect association patterns are characterized by the two items participating in a relationship, along with all other items mediating the interaction (This is explained in detail in Indirect Association: Mining Higher Order Dependencies in Data). For example, Nike and Reebok are indirectly associated as customers seldom buy them together. They tend to be bought separately and hence they are competitors. In a similar way, when searching for investment alternatives, this rule was applied so to empower an algorithmic narrowing down of investment choices.

The results from the algorithm JRip classification rules accuracy for 8 clusters

Having evaluated results for 10 –fold cross-validation with training set size with the above-described methods, 66% show:

- high precision – 0.959
- recall – 0.958
- F1-measure – 0.958

For example, one of the companies for which alternatives were sought, the calculated top 5 investment alternatives were startup companies with the same location, comparable rank,

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

similar number of employees and investors and funding, but from different industries – software, electronics, finance, technologies, merchandise.

This alone supports the main objective of the designed investment recommendation system as it is able to diversify the investment portfolio.

Epilogue: Limitless!?

In her book *Artificial Unintelligence*, author Meredith Broussard suggests that if we understand the limits of what we can do with technology, we can make smarter choices about what we should do with it to make the world better for everyone.

Full automation of investing activities is still far from being possible, one of the many reasons being that algorithms are limited to the quantity and the quality of data they are fed and trained with. However, we, humans, could hardly compete with processing around 8 million companies unarmed. And this is where machines might be of good use for finding investment alternatives. They can help us overcome our natural limitations by leaving the legwork of searching through millions of companies to computers and the art of spotting investment opportunity to investors – where it has always belonged.

Learn how to boost your competitive investment insight!

Book a Discovery Call

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com