

Using Artificial Intelligence and Web Media Data to Evaluate the Growth Potential of Companies in Emerging Industry Sectors

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

“The world needs a better understanding of how to encourage innovation. And innovators need to get better at it. Sign me up.”

Robert Metcalfe

Co-Inventor of Ethernet and Founder of 3Com

In this article, we describe our efforts to adapt and validate a web search and analytics tool – the Gnowit Cognitive Insight Engine – to evaluate the growth and competitive potential of new technology startups and existing firms in the newly emerging precision medicine sector. The results are based on two different search ontologies and two different samples of firms. The first sample includes established drug companies operating in the precision medicine field and was used to estimate the relationship between the firms’ innovativeness and the extent of online discussions focusing on their potential growth. The second sample includes new technology firms in the same sector. The firms in the second sample were used as test cases to determine whether their growth-related web search scores would relate to the degree of their innovativeness. The second part of the study applied the same methodology to the real-time monitoring of the firms’ competitive actions. In our findings, we see that our methodology reveals a moderate degree of correlation between the Insight Engine’s algorithmically computed relevance scores and independent measures of innovation potential. The existence of such correlations invites future work in attempting to analyze company growth potential using techniques founded in web content scraping, natural language processing, and machine learning.

Introduction

Efficient, intelligent tools to evaluate the growth and competitive potential of new technology startups and existing firms in a newly emerging industry sector would deliver valuable insight to stakeholders in many spheres: companies, entrepreneurs, governments at all levels, and investors. However, attempts to study such firms are typically hindered by a lack of sufficient information about the new business sector and the reluctance of the firms – especially startups – to openly share detailed information about their specific business intentions and their actual or potential competitive moves.

In this article, we focus on a sector that typifies these challenges: the emerging precision medicine sector of the healthcare industry. Precision medicine is a new medical approach that promotes the customization of healthcare with medical decisions, practices, and

products that are tailored to specific patient groups or individual patients sharing the same treatment needs (Vicini et al., 2016). The tools employed in precision medicine include molecular diagnostics, imaging, and analytics that are used to select appropriate and optimal therapies based on the context of a patient’s genetic makeup or the results of other molecular or cellular analyses. The establishment of the precision medicine sector has enabled the emergence of new startups focusing on the development of such types of tools as well as the engagement of incumbent medical companies in the exploration of similar business opportunities.

The case of technology startups in this business area is of particular interest. They are expected to win large business deals by competing to create value in the precision medicine value chain through the development of platforms that could be used to interpret and analyze data, thereby contributing to better diagnoses.

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

However, it is very difficult to study such firms due to the insufficiency of information and their own discretion about their specific competitive differentiation strategies. The present article suggests a research approach that could deal with such information insufficiency and discretion by using keyword-based web search techniques and public information that was made available online by third parties. This approach represents an original attempt to study companies' potential for innovation by tuning a developing, innovative competitive intelligence platform for the task of evaluating media perceptions of competitive advantage and growth potential in the precision medicine space.

The article is structured as follows. First, we further describe precision medicine, including both the scope of opportunities in the sector and the challenges of studying the potential growth of companies within it. Then, we describe our research design and the Gnowit Cognitive Insight Engine, which we used for the analyses of the firms' growth potential, innovativeness, and competitive actions. After presenting and discussing the results of these analyses, we offer conclusions and identify future areas of research.

Background: The Precision Medicine Sector

The sphere of precision medicine represents a dramatic departure from prior perspectives on innovation in healthcare. Whereas prior medical treatments have widely been based on targeting an "average patient" in the general population, precision medicine describes an approach in which individuals' measurable unique characteristics across a multitude of genetic, medical, and lifestyle-related parameters define how treatments should be administered (White House, 2015). Instead of rigid prescribed treatment regimes, precision medicine offers highly customizable treatment programmes tailored to provide much more efficient individualized care (Wikipedia, 2017). In 2015, to support innovation in precision medicine, the United States government launched the Precision Medicine Initiative and the affiliated All of Us Research Program (allofus.nih.gov). The All of Us program aims to aggregate data from a population of one million volunteers, including data on genomes, microbiomes, and epigenomes. The data collected by All of Us will be federated in a large database accessible to researchers for scientific purposes.

It is expected that broad effects on healthcare will manifest from developments in precision medicine. Better

understanding and use of individual genetic signals, more customizable medical devices, and personalized diagnoses based on deep analyses of patient data represent some of the foremost aspirations with regard to potential for advancement in this field. As the field develops, new business models will emerge from companies of all sizes in their efforts to bring products to market. These models will be enabled by technological advancement and innovation in spheres including molecular diagnostics, medical imaging, and big data analytics (Knowledge@Wharton & SAP, 2016).

Business researchers and practitioners are increasingly focusing on studying the potential growth of specific companies in the precision medicine sector as well as the innovation strategies that could make a difference in terms of potential patient health outcomes. For example, a recent study provided ranked 23 biopharmaceutical companies based on the extent to which they are driving innovation in precision medicine (Amplion, 2016). The ranking is based on: i) how many precision medicine products a company currently markets; ii) the percentages of the company's recent clinical trials that involve biomarkers; iii) the number of novel oncology drugs the company has in its pipeline; and iv) the number of proprietary biomarkers the company is using in its clinical trials.

Another recent study listed a number of technology companies that are likely to be leaders in precision medicine for years to come (Borukhovich, 2016). Their selection was based on their innovation potential due to their listing in the Crunchbase company database (www.crunchbase.com). The companies are not ranked in any way because they are at a relatively early stage of their lifecycles. The list of companies allows the analysis of emerging value propositions in the precision medicine technology sector. However, the early stage of the companies suggests that many of them are in a stealth mode that does not allow them to share the details of their specific competitive strategies and the analysis of their growth potential. Thus, business analysts, researchers, and potential investors are left to their own means and to information that is available online in examining the growth, innovation, and investment potential of such early-stage firms. The situation highlights the increasing need to develop appropriate competitive intelligence tools and techniques that could use publicly available online information about newly emerging firms to evaluate their growth potential and competitiveness.

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

Research Methodology

The methodology used in the research for this article was structured around using keyword-based web search and real-time media monitoring techniques to analyze publicly available online data about companies to be studied. This approach was taken in combination with the use of the Gnowit Cognitive Insight Engine (Gnowit engine, www.gnowit.com), a platform that uses machine learning to analyze the magnitude of similarity between corpuses of documents. The Gnowit engine was principally developed by Andrew D. Droll, Lead Data Scientist at Gnowit Inc., by building on the expertise of Shahzad Khan, CTO of Gnowit Inc.; together, they are the first two co-authors of this article. The research methodology was shaped in collaboration between Gnowit and the two other co-authors who are researchers at the University of Southern Denmark. There are at least two substantial prospective advantages associated with focusing on information that is provided online by third parties. First, in many cases, this is the only valuable information that is available about early-stage companies. Second, information provided by third parties is expected to be unbiased with respect to the scarce marketing information provided by the companies themselves. With this research, we aimed to determine whether this type of data, when combined with a machine learning approach, could produce useful insight into companies' growth potential.

We started our study by examining the relationship between the intensity of online media discussions focusing on the business growth of companies in the pharmaceutical and biotechnology industries, and the empirical ranking of their innovativeness such as provided by independent third-party sources. To this end, the Gnowit engine was employed to analyze textual data collected from online publications that are available throughout the Internet. Gnowit's crawler architecture automatically collects news article content from more than 8,000 publications, including the largest newspapers in the world, and extending down to a number of small regional and local papers and outlets. All of this data was available to the Gnowit engine during our experimentation.

To analyze this aggregated data using the Gnowit engine, we have developed a Boolean query designed to extract articles relating to industry growth in the precision medicine sector. The specific keywords were selected through a close examination of recent books such as *Get Backed* (Baehr & Loomis, 2015), *The Lean Startup* (Erik Ries, 2011), and *Business Model Generation* (Oster-

walder & Pigneur, 2010). The final query combined a variety of composite keywords focusing on:

Market growth ("Market size" OR "market growth" OR ((fast OR growing) AND market) OR "market conditions" OR "attractive market position" OR "unique market position" OR "emerging market" OR "niche market" OR "new market" OR "market confusion" OR "market ambiguity" OR "market share" OR "market leader");

Business growth ("Growth potential" OR "business growth" OR "growing momentum" OR "growth strategy" OR "fast-growing company" OR "growing sales" OR "sales growth" OR "revenue growth" OR "ramp up sales" OR "growing market demand" OR "growing customer demand" OR "scale up" OR "boosting revenue" OR "increasing revenue" OR "greater product sales" OR "new service income" OR "minimize costs" OR "new revenue streams" OR "sustainable revenue" OR "stimulate growth" OR "growth indicators" OR "growth factors" OR "competitor growth strategy" OR "competitive strengths" OR "business strengths" OR "increasing return" OR "increasing income");

Competitive differentiation strategy ("Business differentiation" OR "competitive differentiation" OR "market differentiation" OR "competitive advantage" OR "business advantage" OR "unique value proposition" OR "unique selling point" OR USP OR "competitive position").

The three composite keywords above were connected through a logical OR operator. The resulting larger composite keyword was connected through a logical AND with the terms "precision medicine" and "personalized medicine". The final result was a query, or "search ontology", which we labelled Precision Medicine Growth (PMG). The Gnowit engine is usually used to compare companies on the basis of their relationship to certain specific single signals; however, Gnowit's systems permit comparisons across several signals at once, meaning it can classify a company's media coverage in a true ontological fashion. This query served as a heuristic to isolate a set from several document sources that were initially pre-collected and maintained by Gnowit: Canadian Online News, Government – Med Sources, Tech Startup News – USA, and US Online News.

The Gnowit engine used the PMG query as an instruction to perform the following steps:

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

1. Retrieve the collection of the latest published articles that match the query heuristic.
2. Tokenize the content of these articles in a way that leads to the construction of a set of textual terms and numeric weightings for each term. This step transforms each document into a term-weighting vector.
3. Add the individual document term-weighting vectors together to create an aggregate term-weighting vector for the entire PMG heuristic (henceforth called the “PMG vector”).
4. Store this aggregated heuristic vector as a retrievable context vector that can be used to provide numeric strength-of-similarity analysis with other similar heuristically-developed aggregate vectors.

The capability described above was leveraged as follows. We used as a case study the sample of 23 companies that were ranked by (Amplion, 2016) in terms of their innovativeness in precision medicine. For each of these companies, we developed an additional individual Boolean query designed to act as a heuristic returning articles that relate to that company name alone. Subsequently, we submit an instruction to the GCIE that requests the computation of an aggregate heuristic vector (in the same way as described above for generating the PMG vector) for each company name heuristic, and returns back a numeric value corresponding to its strength-of-correspondence to the initially calculated PMG vector. As a result, each company now has a numerical strength-of-correspondence measure that could be used to evaluate its potential growth in the precision medicine sector. We refer to this measure as the PMG Gnowit score.

In the second part of our work, we applied the above methodology to another research context – real-time monitoring of the competitive actions of the same companies. This specific choice of the second research context was motivated by the insights from a closer examination of the way Gnowit search engine operates. The Gnowit engine does not search across all articles on the Internet; rather, it only searches across those articles that were published in a period starting two weeks in the past and ending at the present moment – a limitation imposed by the company’s infrastructure cost constraints as a small startup. The time period for the search is a moving time window that looks only two weeks back in time. In this sense, it might be best suited for the monitoring of most recent events or actions corresponding to firms’ current competitive strategy. That

is why we constructed a new search ontology focusing on competitive actions. The construction of the new ontology was done on the basis of insights from McInnis (2008) and Grimm and colleagues (2006). The query corresponding to competitive actions in precision medicine (CAPM) took the following form:

((“Asset development” OR “Resource adjustment”) OR (“Brand awareness” OR “Brand loyalty”) OR (“Intellectual property” OR “Product Development” OR “Product improvement” OR “Product design”) OR (“Competitive strategy” OR “Competitive advantage” OR “Competitive market position” OR “strong market position” OR “Time to market” OR “lower entry barriers” OR “Joint ventures” OR “Economies of scale” OR “economy of scale”) OR (“accelerate adoption” OR “market adoption” OR “viable business model” OR “solid business model” OR “category leader” OR “customer engagement” OR “customer involvement” OR “User driven”) OR (“Demonstrate performance” OR “Development center” OR “Exceed goals” OR “Expand adoption” OR “Deliver more value”) OR (Funding OR “Global expansion” OR “Global reach” OR Grow OR growth OR “product availability” OR “Strength management” OR “technology support”) OR (Innovation OR “innovation at a fraction of the cost” OR “cost effective innovation” OR “innovation at lower cost” OR “Product innovation”) OR (“join executive” OR “join leaders” OR “joint solution”) OR (“launch product” OR “launch solution” OR “lead emerging technology” OR “lock-in” OR “lower cost ownership”) OR (“new alternative” OR partner OR “partner program” OR “partner trusted provider” OR “powerful capabilities” OR “rapid development” OR “rapid innovation” OR “reduce costs” OR “regional expansion”)) AND (“precision medicine” OR “personalized medicine”)

The replication of the search methodology on the basis of the CAPM ontology resulted in a CAPM vector that was used to measure the competitive activity of PM firms. We refer to this measure as the CAPM Gnowit score.

Research Findings

The methodology described in the previous section allowed us to evaluate the correlation between Amplion’s innovation ranking and the PMG Gnowit score, which corresponds to the intensity of online media discussions of companies’ focus growth in the precision medicine sector.

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

Ranking firms by PMG Gnowit scores

Table 1 lists the 23 companies that were ranked by Amplion (2016) in terms of their innovativeness in the precision medicine field. The table also includes their innovation ranking and two sets of PMG Gnowit scores corresponding to their growth potential in the precision medicine sector. The results in Table 1 compare the firms' innovation ranking and the PMG Gnowit scores for two different two-week time periods: December 13 to 17, 2016, and January 20 to February 3, 2017, 2017. Figure 1 illustrates the PMG Gnowit scores for all 23 firms in the second time period. The numerical difference in scores between these periods is due to shifting coverage of these companies and of the precision medicine space over time. Variance of this type is normal because it is based on current news coverage. The main element of interest is that of scores relative to each other across different companies (though this will also shift as news coverage changes, for example in response to product announcements or releases or company financial announcements).

Figure 1 shows that the highest scores in the second period belong to Johnson & Johnson, followed by Merck, Pfizer, Bristol-Myers Squibb, and Roche. Table 1 also reveals substantial differences in the scores between successive time periods depending on the degree of online discussions associated with a particular firm. For example, there was a 10-times increase in Roche's PMG Gnowit score in the second time period. Such radical changes could be explained by the nature of the Gnowit search engine, which focuses on monitoring online discussions that happened in the last two-week period. It could happen then, for example in a time of reporting impressive (or disappointing) quarter results, that the name of a specific company has been

involved in an unusually high number of online discussions in the past two weeks, leading to an unusually high PMG Gnowit score.

The above discussion leads to three important points:

1. The suitability of the Gnowit engine for a specific search will depend on the context of the research project. In our present case, we are trying to evaluate the growth potential of firms in an emerging industry sector. It is clear that the evaluation of such potential should not be based on online information collected over just two weeks. The evaluation of the growth potential should extend the methodology by including the possibility to accumulate or integrate results from multiple successive two-weeks periods in order to cover a longer period of time, such as six months, a year, or longer.
2. The evaluation of the growth potential cannot be based on the quantitative results alone and should incorporate additional qualitative semantic textual analysis. The Gnowit search platform offers additional analytical capabilities that allow for sentiment analysis to indicate the degree to which the online discussions are associated with positive or negative statements. This feature was not included in the present study and will become the subject of future studies.
3. The two-week search window feature of the Gnowit search engine suggests that it could be more directly applicable to the real-time monitoring of firms' competitive actions because such actions might have a real-time competitive effect on competitors.

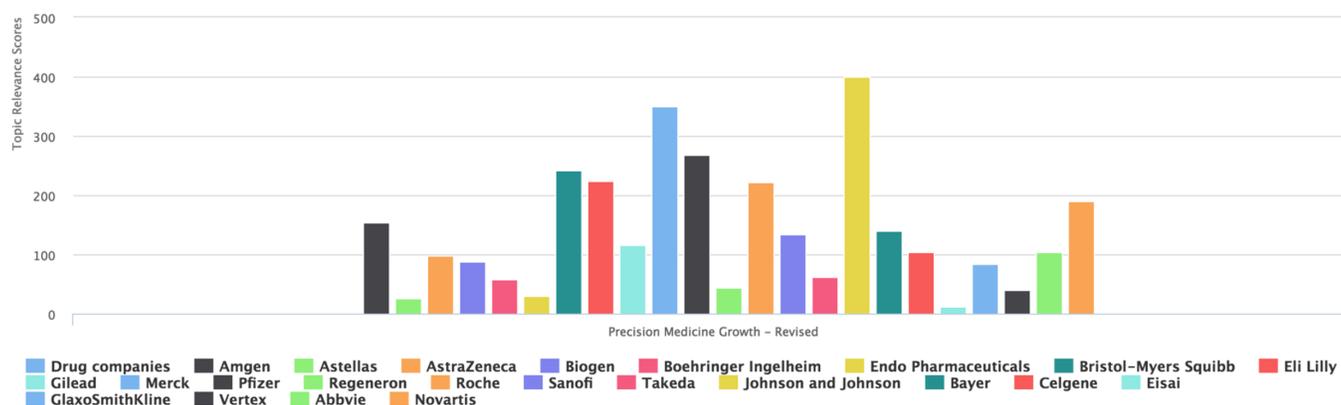


Figure 1. Ranking of the 23 innovative drug companies in terms of their PMG Gnowit scores. The graph was automatically generated by the Gnowit search platform.

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

Table 1. Innovation ranking of 23 established firms in the emerging precision medicine sector (Amplion, 2016) together with their PMG Gnowit scores corresponding to the intensity of online discussions about their growth potential in the PM sector

No.	Company Name	Innovation Ranking	PMG Gnowit Score Period 1	PMG Gnowit Score Period 2
1	Novartis	0.67	23.11	190.76
2	Roche	0.56	20.97	222.49
3	AstraZeneca	0.46	11.99	98.69
4	Pfizer	0.43	30.18	268.36
5	GlaxoSmithKline	0.38	6.54	83.73
6	Boehringer Ingelheim	0.35	6.74	57.82
7	Merck	0.31	21.46	349.55
8	Celgene	0.30	13.02	103.31
9	Eisai	0.29	2.23	11.11
10	Johnson & Johnson	0.25	24.89	410.11
11	Sanofi	0.21	10.94	134.62
12	Lilly	0.18	21.08	224.97
13	Bristol-Myers Squibb	0.18	33.59	245.19
14	Bayer	0.15	13.98	140.34
15	Amgen	0.14	9.40	154.13
16	Astellas	0.13	4.26	25.70
17	Abbvie	0.13	13.08	104.33
18	Takeda	0.11	3.02	61.55
19	Gilead	0.10	19.94	116.29
20	Endo	0.09	7.77	29.41
21	Biogen	0.06	16.15	88.25
22	Vertex	0.03	4.05	40.08
23	Regeneron	0.02	5.35	43.22

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

Relationship between innovativeness and PMG Gnowit score

For the first measurement period, the Spearman correlation coefficient between the innovation ranking and the PMG Gnowit score was 0.38 ($p < 0.1$). In the second period, it was 0.43 ($p < 0.05$). Even though the statistically significant correlation coefficients correspond to a moderate degree of correlation, their degree of statistical significance may vary across different time periods depending on the intensity of online discussions in a specific time period. Nevertheless, a moderate or high statistically significant degree of correlation suggests the existence of a relationship between a firm's PMG Gnowit score and its innovativeness. Such relationship will obviously depend on the specific terminological composition of the search ontology. We can claim, however, that our growth potential search ontology seems to have worked well in detecting a relationship between growth terminology and innovativeness. The existence of such relationship suggests that the methodology based on our growth potential search ontology could be applied to study the growth and innovation potential of early-stage companies for which there is no competitive intelligence information. To test this assumption, we will apply the methodology to the precision medicine technology companies that were identified as promising by Borukhovich (2016).

Applying the methodology to studying the innovativeness of new precision medicine technology firms

In this section, we present results of the PMG Gnowit score ranking of a sample of 29 new technology-based firms in the precision medicine sector (Borukhovich, 2016). Table 2 shows the results for the PMG Gnowit scores of these firms in the second time period. Only 6 firms had a nonzero score (Table and Figure 2). The rest of the firms did not have any online coverage in this particular period.

The previous results about the correlation between firms' innovativeness and their PMG Gnowit scores, together with the results shown in Table 2, were expected to provide us with a basis to estimate the degree of innovativeness of the firms shown in Figure 2. In addition, the results shown in Table 2 would suggest that there was no innovative activity by most of the firms. Such conclusions, however, should be subjected to the same considerations as the ones made in the previous sections. First, the results as such do not allow such conclusions unless we have performed online measurements across multiple successive two-week periods, which would allow us to estimate the aggregated effect of the online discussions over a longer period of time. Second,

Table 2. PMG Gnowit scores of 29 innovative new technology firms in the precision medicine sector

No.	Company Name	PMG Gnowit Score
1	KineMed	0
2	Advanced Cell Diagnostics	0.43
3	Syapse	0
4	Centrillion Biosciences	0
5	Epic Sciences	0.32
6	Ignyta	1.77
7	Eve Biomedical	0
8	Sctheranostics	0
9	Kuraoncology	0
10	Genoptix	1.07
11	Precipio Diagnostics	0
12	Inform Genomics	0
13	Miramix	0
14	Enumeral Biomedical	0
15	Population Diagnostics	1.08
16	Genterpret	0
17	Rosetta Genomics	0
18	Translational Software	0
19	4D Healthcare	0
20	Everist Health	0
21	GenomOncology	0
22	Cellular Dynamics Int	7.60
23	Kailos Genetics	0
24	Microarrays	0
25	PHIGENIX	0
26	Precision for Medicine	0
27	NeoDiagnostix	0
28	Genomind	0
29	Permedly	0

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

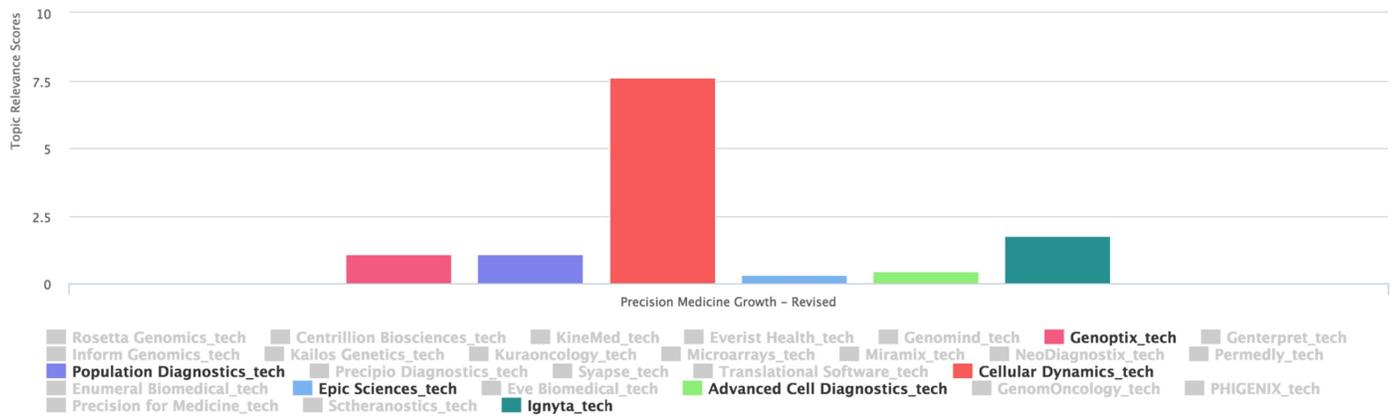


Figure 2. Ranking of the six new technology firms with growth-oriented online coverage in the second time period

the functionality of the Gnowit search engine could be better suited to search ontologies focusing on monitoring real-time competitive moves of firms in an emerging industry sector.

Applying the methodology to studying the competitive actions of precision medicine firms

In this section, we apply the competitive action search ontology to study the real-time competitive actions of the same two samples of firms that were studied in the previous sections. Table 3 shows the list of the firms together with their innovation ranking (the same as Table 1) and their CAPM Gnowit scores. Figure 3 provides ranks the firms in terms of their CAPM Gnowit scores (measuring coverage related to competitive advantage in the precision medicine space). The correlation coefficient between firms' innovativeness and their CAPM Gnowit scores is 0.439 ($p < 0.05$), which suggests again a moderate degree of correlation.

The visible degree of correlation between the firms' degree of innovativeness and their CAPM Gnowit score again offers the opportunity to use the competitive action ontology to estimate the innovation potential of the 29 new technology firms in the precision medicine sector (Borukhovich, 2016).

Table 4 lists 5 out of the 29 firms that had some online coverage in the second time period; the rest of the firms did not have such coverage. Figure 4 shows the five firms in terms of their CAPM Gnowit scores. The results show that Cellular Dynamics Int (CDI) has the highest CAPM Gnowit score. Table 5 provides details of three documents referring to CDI that were found to be most relevant in the search. The relevance of a particular online document within the context of the specific search project is defined by a relevance score calculated by the Gnowit search platform on the basis of the semantic correlation of the content of the document and the

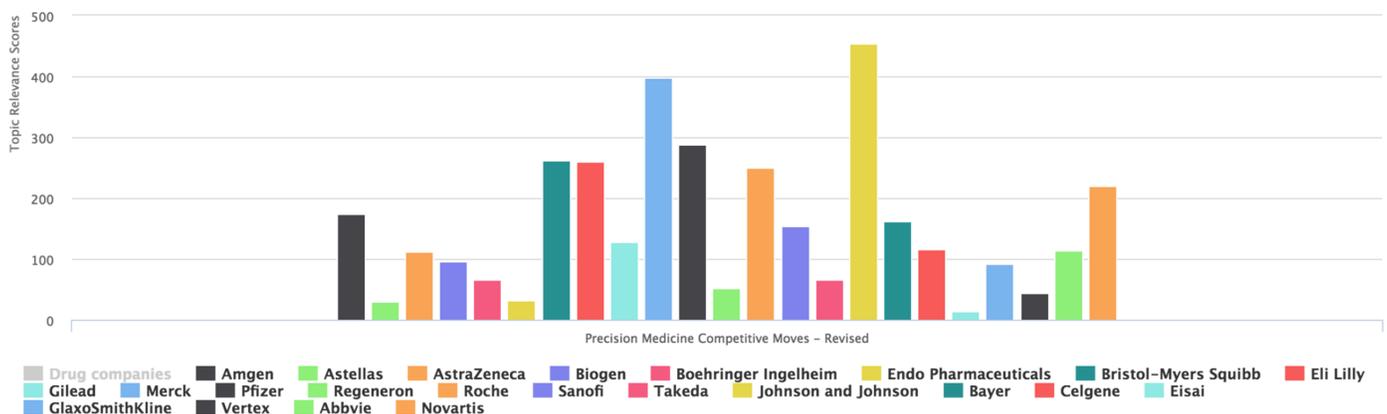


Figure 3. CAPM Gnowit score ranking of the 23 drug firms operating in the precision medicine sector over the second time period

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

Table 3. CAPM Gnowit scores of the 23 drug firms over the second time period

No.	Company Name	Innovation Ranking	CAPM Gnowit Score
1	Novartis	0.67	219.75
2	Roche	0.56	250.44
3	AstraZeneca	0.46	111.12
4	Pfizer	0.43	288.43
5	GlaxoSmithKline	0.38	91.83
6	Boehringer Ingelheim	0.35	65.84
7	Merck	0.31	398.48
8	Celgene	0.30	116.32
9	Eisai	0.29	13.16
10	Johnson & Johnson	0.25	453.22
11	Sanofi	0.21	153.83
12	Lilly	0.18	260.89
13	Bristol-Myers Squibb	0.18	261.63
14	Bayer	0.15	161.82
15	Amgen	0.14	174.08
16	Astellas	0.13	29.78
17	Abbvie	0.13	113.93
18	Takeda	0.11	66.28
19	Gilead	0.10	127.46
20	Endo	0.09	32.90
21	Biogen	0.06	95.72
22	Vertex	0.03	43.62
23	Regeneron	0.02	52.53

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

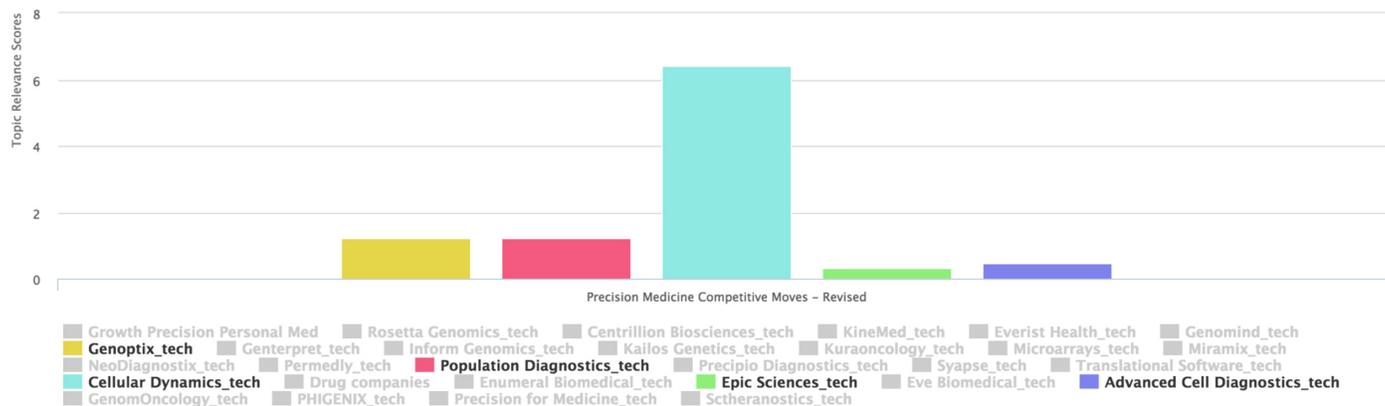


Figure 4. CAPM Gnowit score ranking of five new technology firms operating in precision medicine sector

Table 4. CAPM Gnowit scores of five innovative new technology firms in the precision medicine sector that had online coverage focusing on competitive actions

No.	Company Name	CAPM Gnowit Score
1	Cellular Dynamics Int (CDI)	6.41
2	Genoptix	1.23
3	Population Diagnostic	1.23
4	Advanced Cell Diagnostics	0.46
5	Epic Sciences	0.34

Table 5. List of the online documents about Cellular Dynamics Int (CDI) with the highest relevance for the search focusing on competitive actions

No.	CDI-Related Document	Relevance Score and Web Link	Context
1	2015 FDA Science Forum Emerging Technologies, May 27 – 28, 2015 U.S., Food and Drug Administration	Document relevance score: 100 Web link: tinyurl.com/yatn6oy2	CDI used by United States Food and Drug Administration (FDA) scholars as a reference in a comparison with another major drug supplier
2	Global Stem Cell Therapy Applications & Treatment Analysis & Trends 2016-2025: \$1.55 Billion Growth Opportunities/ Investment Opportunities – Research and Markets	Document relevance score: 62 Web link: tinyurl.com/y8tdnu13	CDI included in the list of leading companies in stem cell therapy and treatment analysis
3	FDA Cooperative Research and Development Agreements (CRADAs)	Document relevance score: 7 Web link: tinyurl.com/yc8464ja	CDI project listed as one of the projects supported by the FDA R&D program

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

search ontology. The specifics of the semantic correlation are out of the scope of this article; however, for a high-level view, the scores are derived in the following way. First, Gnowit's search engine is queried for the articles matching the search ontology terms. These articles are analyzed using natural language processing techniques to extract a characteristic linguistic profile for the associated query in recent media coverage. Then, when plots are generated, the same process is performed for companies themselves. Articles relating to each company are retrieved from Gnowit's search engine via query, and these are analyzed to produce a characteristic linguistic profile for the companies. Finally, a similarity measure is applied to generate linguistic proximity scores between media coverage relating to the search ontology terms and the companies undergoing analysis. By comparing these scores across companies using the same underlying search ontology terms, we obtain relative measures of company discussion in the media.

The information provided in Table 5 offers just a glance of how the Gnowit search engine could be used to monitor the competitive actions of new firms in an emerging industry sectors. The validation and automation of such search projects needs more systematic research studies, which will be the subject of future efforts.

Conclusion

This article presented the results of a research study focusing on adapting and validating the use of a business intelligence tool – the Gnowit Cognitive Insight Engine – for entrepreneurship and innovation research. The focus of the study was on uncovering the exploratory and analytics potential of the existing functionality of the Gnowit search engine and the ultimate objective is to articulate insights that could further enhance its potential as a competitive intelligence tool for entrepreneurs, investors, managers, and entrepreneurship scholars. The results provide a first glance at this potential and highlight opportunities for future studies in this direction.

In particular, we hope in the future to be able to expand the time interval for news coverage to be analyzed beyond Gnowit's current two-week scope. A longer time period will allow more comprehensive and significant evaluations of the significance of the types of measures that we examine in this article. Also, this avenue could open the possibility of developing an engine to monitor changes in companies' potential over time, potentially kept up-to-date in real time, which could provide true utility to researchers and companies in the wild. As the current time restriction is primarily one borne of resource constraints, investments in Gnowit's technology may permit this type of exploration in the future. Finally, additional research opportunities are presented by the capabilities of Gnowit's technology not yet explored by this article. In particular, in this article, we only examined comparisons between companies using search ontologies consisting of a single search term – in effect, doing one-dimensional analyses of the companies' relative media impressions. The Gnowit ontology engine supports ontologies with dozens of terms, which might permit the generation of significant, multi-dimensional comparisons and analyses of companies' impressions in the media across many signals of interest.

Acknowledgements

An earlier version of this article was presented at the ISPIM Innovation Forum in Toronto, Canada, March 19–22, 2017. ISPIM (ispim-innovation.com) – the International Society for Professional Innovation Management – is a network of researchers, industrialists, consultants, and public bodies who share an interest in innovation management.

The authors express their gratitude to Dr. Vlad Valtchinov, Principal, New Atlantic Technology Group, LLC, Boston, USA, for the fruitful discussions that helped to identify the emerging precision medicine sector as an appropriate research area.

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

About the Authors

Andrew Droll is Lead Data Scientist at Gnowit in Ottawa, Canada. Andrew holds PhD and MSc degrees in pure mathematics from Queen's University in Kingston, Canada, and he holds a BSc degree in Mathematics and Physics from Carleton University in Ottawa, Canada. His peer-reviewed publications span the fields of physics, mathematics, and computer science. Currently, Andrew works on development and management of Gnowit's research and engineering projects.

Shahzad Khan is the CTO of Gnowit Inc. in Ottawa, Canada, that provides personalized, real-time web intelligence for individuals and corporations. The firm employs artificial intelligence to automatically gather data from fragmented web sources in near-real-time and filter the data using human-like synthetic cognitive methods to provide highly curated intelligence to their clients. He has a PhD in Computer Science from the University of Cambridge, United Kingdom, an MSc in Information Studies from Syracuse University in New York, USA, and a BSc (Hons) in Computer Science from the Lahore University of Management Science (LUMS) in Lahore, Pakistan. His research interests lie in semantic analysis on big data repositories using natural language processing and machine learning at scale.

Ehsan Ekhlās is a student and entrepreneur completing studies in Technology Entrepreneurship and Business Innovation at the University of Southern Denmark. Ehsan is also Founder & CEO of Mamac IVS, a company focused on fashion accessories for Apple products. In his research, Ehsan uses technological and big data tools to try to discover insights about how people do work in the real world.

Stoyan Tanev is an Associate Professor in the Department of Technology and Innovation, Faculty of Engineering, Southern Denmark University (SDU) in Odense. Dr. Tanev is leading the Technology Entrepreneurship stream of the Master Program of Product Development and Innovation at SDU. He is also an Adjunct Research Professor in the Sprott School of Business at Carleton University in Ottawa, Canada, where he is associated with the Technology Innovation Management Program. He has a MSc and a PhD in Physics jointly from the University Pierre and Marie Curie, Paris, France, and the University of Sofia, Bulgaria, a PhD in Theology from the University of Sofia, Bulgaria, an MEng in Technology Innovation Management from Carleton University, Canada, and an MA from the University of Sherbrooke, Canada. He has multidisciplinary research interests with a focus on the fields of global technology entrepreneurship, technology innovation management, business model design, and value co-creation. Dr. Tanev is Senior IEEE member, as well as member of the editorial boards of the *Technology Innovation Management Review*, the *International Journal of Actor-Network Theory*, and *Technological Innovation*.

References

- Amplion. 2016. Precision Medicine Scorecard: Drug Companies, Q1 2016. *Amplion.com*, April 28, 2016. Accessed June 1, 2017: <http://www.amplion.com/download-the-precision-medicine-scorecard-q1-2016>
- Baehr, E., & Loomis, E. 2015. *Get Backed: Craft Your Story, Build the Perfect Pitch Deck, and Launch the Venture of Your Dreams*. Boston, MA: Harvard Business Review Press.
- Borukhovich, E. 2016. Tech and Precision Medicine: 29 Companies Who Will Lead the Charge. *HealthWorksCollective*, January 22, 2016. Accessed June 1, 2017: <https://www.healthworkscollective.com/tech-and-precision-medicine-29-companies-who-will-lead-charge/>
- Grimm, C., Lee, H., & Smith, K. (Eds.). 2006. *Strategy as Action: Competitive Dynamics and Competitive Advantage*. Oxford: Oxford University Press.
- McInnis, G. 2008. *Competitive Actions of Companies Whose Revenue Relies on Open Source Software*. Master of Applied Science in Technology Innovation Management Thesis, Carleton University, Ottawa, Canada.
- Osterwalder, A., & Pigneur, Y. 2010. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Chichester, UK: John Wiley & Sons.
- Ries, E. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. New York: Crown Business.

Using AI and Web Media Data to Evaluate the Growth Potential of Companies

Andrew Droll, Shahzad Khan, Ehsanullah Ekhlās, and Stoyan Tanev

Tanev, S., Liotta, G., & Kleismantas. 2015. A Business Intelligence Approach Using Web Search Tools and Online Data Reduction Techniques to Examine the Value of Product-Enabled Services. *Expert Systems with Applications*, 42(21): 7582–7600. <https://doi.org/10.1016/j.eswa.2015.06.006>

Vicini, P., Fields, O., Lai, E., Litwack, E. D., Martin, A. M., Morgan, T. M., Pacanowski, M. A., Papaluca, M., Perez, O. D., Ringel, M. S., Robson, M., Sakul, H., Vockley, J., Zaks, T., Dolsten, M., & Søgaard, M. 2016. Precision Medicine in the Age of Big Data: The Present and Future Role of Large-Scale Unbiased Sequencing in Drug Discovery and Development. *Clinical Pharmacology & Therapeutics*, 99: 198–207. <https://doi.org/10.1002/cpt.293>

White House. 2015. *Precision Medicine Initiative*. Washington, DC: The White House. <https://obamawhitehouse.archives.gov/node/333101>

Wikipedia. 2017. Precision Medicine. *Wikipedia*. Accessed June 22, 2017: https://en.wikipedia.org/wiki/Precision_medicine

Knowledge@Wharton & SAP. 2016. Precision Medicine: New Paradigms, Risks and Opportunities. *Knowledge@Wharton*, September 22, 2016. Retrieved on June 23, 2017: <http://knowledge.wharton.upenn.edu/article/precision-medicine-new-paradigms-risks-opportunities/>

Citation: Droll, A., Khan, S., Ekhlās, E., Tanev, S. 2017. Using Artificial Intelligence and Web Media Data to Evaluate the Growth Potential of Companies in Emerging Industry Sectors. *Technology Innovation Management Review*, 7(6): 25–37. <http://timreview.ca/article/1082>



Keywords: startup growth potential, artificial intelligence, entrepreneurship, precision medicine sector, online textual data, analytics, business intelligence