



Northern Light Whitepaper

Search Technology for Strategic Research Portals

Northern Light®

March 2013

TABLE OF CONTENTS

Background On Northern Light and SinglePoint.....	3
Limitations of Standard Enterprise Search Technology For Research Portals.....	2
Lowest Common Denominator Search.....	2
Enterprise Search and the Problem of External Content.....	3
External Content Aggregation.....	3
External Content Access Rights.....	3
Copyright Compliance.....	4
Aggregation Of Licensed Content Onto the Enterprise Network.....	4
Federated Search Versus Integrated Search.....	5
Northern Light Search and Search-Related Services.....	9
Third-Party External Content Aggregation.....	11
Provide Content Aggregation As a Key Service.....	11
Aggregate and Search Syndicated Research Analyst Content.....	12
Support Third-Party Subscription Access Models.....	13
Enforce Third-Party Copyright Restrictions.....	14
Business News.....	15
Northern Light Business News.....	15
Licensed News.....	16
Scientific and Academic Journals.....	16
Life Sciences Conference Presentation Abstracts.....	16
Social Media.....	16
Custom Web Search.....	16
Government Databases and REgulatory Information.....	16
Financial Analyst Reports and Public Financial Data.....	17
Structured Data From Data Vendors.....	17
Video Content.....	17
Internal document repositories of primary research and market intelligence.....	17
Custom Content Aggregation.....	18
Northern Light Search Engine.....	19

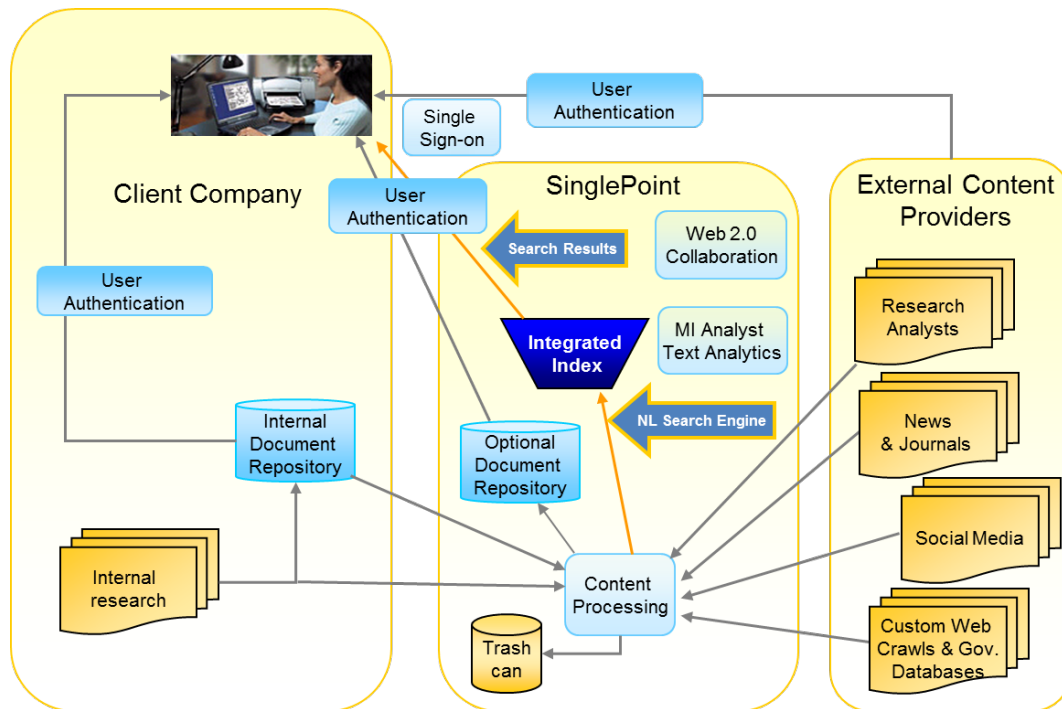
Stemmed Index.....	19
Unstemmed Index and Literal Searching.....	20
Proximity Index and Proximity Searching.....	20
Phrase Searching	21
Concept Index and Concept Searching	21
Name Index and Name Searching.....	21
Acronym Index and Acronym Searching.....	22
Relevance Ranking Optimized For Business and Technology Research	22
Customizable Relevance Ranking	26
Extensive Query Syntax Options.....	27
Complex Boolean Expressions	27
Wildcards.....	28
Metadata Indexing and Search.....	28
Combined Syntax.....	29
Really Large Queries	30
Thesaurus Expansion	31
Faceted Navigation	32
Interoperability With Microsoft SharePoint and Other Enterprise systems	33
MI Analyst – Text Analytics For Industry and Technology Research	34
The Problem	34
Introducing Meaning Extraction	35
Meaning-Loaded Entities.....	35
Concept Identification, Navigation, and Search	36
Automatic Identification of Scenarios of Meaning.....	38
Summary of MI analyst.....	40
In Conclusion	41
Contact Information	43

BACKGROUND ON NORTHERN LIGHT AND SINGLEPOINT

Northern Light is in a single business: providing strategic research portals to global, new product, new technology, and innovation driven organizations in many industries. Our strategic research portals have been used for market research, competitive intelligence, product development, and technology research to global enterprises since 1999. Our current strategic research portal client list reads like a who's who in technology and industry, including leaders in many industries:

- Life sciences
- Financial services
- Telecom
- Information Technology
- Manufacturing
- Agribusiness
- Consumer products
- Logistics and distribution
- Corporate strategy consulting
- Energy

Altogether, there are over 200,000 individual users of our *SinglePoint™* strategic research portals at companies like the ones above. *SinglePoint* is a hosted, turnkey solution provided in its entirety by Northern Light. The client licenses, creates, or simply identifies the content to be included in the research portal, but after that Northern Light handles all other aspects of the portal including development, configuration, deployment, content aggregation, indexing and search, text analysis, collaboration, user management, document security, and reporting. Below is an overview of the *SinglePoint* solution.



A typical *SinglePoint* has 10-20 licensed external syndicated research sources with hundreds of thousands of research reports, several internal primary research repositories, a business news feed, and is used by 5,000 users within our client's organization. Northern Light also harvests content from five thousand websites accounting for 20 million news stories, and indexes 30 million scientific and journal articles.

Each *SinglePoint* client has a custom search index built from the specific content providers and content services that the client licenses and/or wishes to have in the client's strategic research portal. If a client needs a third-party research provider that is not in our universe of existing relationships, Northern Light has an established process for bringing new sources into the *SinglePoint* system. Clients will never outgrow Northern Light's ability to integrate content on the clients' behalf.

Northern Light *SinglePoint* portals enjoy a high ROI, paying for themselves ten to twenty times per year. Northern Light publishes a whitepaper on this topic titled *SinglePoint Strategic Research Portal ROI Analysis* and it is available in the Knowledge Center on our corporate website at <http://northernlight.com/knowledgecenter>.

LIMITATIONS OF STANDARD ENTERPRISE SEARCH TECHNOLOGY FOR RESEARCH PORTALS

There are many enterprise search solutions available on the market from industry leaders such as Oracle, Google (their enterprise search appliance), HP, IBM, open source project Lucene, and Microsoft. Microsoft provides SharePoint Search, previously called FAST, and this is the most common search technology in the enterprise space because it is built into Microsoft SharePoint which is a ubiquitous portal intranet portal platform.

In general, the enterprise search solutions provided by Microsoft and other enterprise search providers share these characteristics:

- Can index and search any common enterprise document type such as MS Office, email, PDF's, and HTML.
- Can scale to high document counts typical of enterprise repositories.
- Have passable relevance ranking for general purpose search of various enterprise documents like email, SharePoint sites, MS Word and PowerPoint, chat logs, intranet pages, etc.
- Have extensive enterprise integration admin tools for doing things like identifying repositories, integrating with Active Directory for document security, etc.
- Support industry standards for interoperability like OpenSearch, OpenAuth, Web Services, ODBC, and Microsoft SharePoint federated search.

LOWEST COMMON DENOMINATOR SEARCH

The problems for enterprise search engines are founded in the need to provide a search solution that works in a wide variety of settings. The reality is that the typical enterprise-wide search engine cannot know the use case of a particular user ahead of time. One user might be searching for the latest expense report forms, another user might be searching for a technology whitepaper, another for the company softball league schedule, and another for a strategic report on marketing trends. The inability to predict a particular user's context and the source, document repository, taxonomy, metadata, search form, and indexing strategy that would serve that particular user best greatly limits the ability of generic horizontal enterprise search to produce superior results relative to the case where these things can be known.

Also, another limiting factor is the broad range of users served and the unknown search skill of a particular user. Because a particular user might have very limited search skills, the typical enterprise search application is optimized around simple search terms, no use of metadata, and no use of advance query syntax. Not having these available as a default option to the professional power researcher impacts those power users negatively.

The requirement to serve all use cases and all comers means the typical deployed enterprise search solution strives to produce a passable result for an astonishingly wide breadth of search use cases, rather than produce a stellar result for a known, single use case. All search settings and defaults are optimized for this goal, producing *lowest common denominator* search. The typical enterprise search application simply cannot be optimized for a particular use case.

ENTERPRISE SEARCH AND THE PROBLEM OF EXTERNAL CONTENT

Another crucial factor deployed enterprise search engines *assume enterprise IT control of the document repositories*. In a related way, deployed enterprise search engines *assume that knowledge about access rights to documents resides in the company's Active Directory* (or in related and accessible places like access control lists in SharePoint sites).

For strategic research applications where the content typically comes in whole or in part from outside the organization, both of these assumptions are invalid.

EXTERNAL CONTENT AGGREGATION

Enterprise search solutions are built on the assumption that the repositories are on the corporate intranet so that content aggregation and integration are not problems that the search solution need address beyond being pointed at the repositories on the enterprise network. The IT department ultimately controls the repository and publishing systems, document formats, metadata, publishing schedules, change process, and other aspects of these repositories.

In a related way, deployed enterprise search engines *assume that knowledge about rights to documents resides on the company's Active Directory* (or in related and accessible places like access control lists in SharePoint sites). IT need do nothing but set the standard parameters for a document repository or file directory and security is automatically taken care off by the Microsoft-based enterprise network processes.

Third-party publishers of strategic research have a myriad of publishing systems, document formats, metadata conventions, publishing schedules, and technical access gateways. No part of the process is the same between two content providers. IT departments within most enterprises have little experience with content sources they have no control over and they are not typically skilled or culturally suited for dealing with the dizzying array of issues that outside publishers of content will present to them.

EXTERNAL CONTENT ACCESS RIGHTS

Third-party publishers of research content also have many business models for licensing their content and typical enterprise security systems have no constructs that match the licensing plans and no means for capturing the

licensing details. For example, a publisher might license on a seat-by-seat, by department, by geographic region, on a role basis, using a bucket-of-reports model, or by a combination of these. Or the publisher might license some of its content using one model, perhaps enterprise-wide, while other parts of their content is licensed using another model, perhaps by the seat. Two employees may sit next to each other, be in the same department, have the same title and rank, do the same work, and one may have access to a particular piece of licensed content in the repository and the other not have any rights to it based on nothing other than the licensing arrangement with the outside party. Even the boss of the authorized employee will often not be an authorized user of seat-limited content. (Indeed, the CEO of the company may not have access if no seat has been purchased for him or her!)

Worse than the complexity of the access rules, *the knowledge of the access rules does not reside on the IT network of the enterprise, it resides on the IT networks of the content providers*. As such it is not available to the typical enterprise search solution at run-time. In a similar vein, the typical enterprise search solution does not have mechanisms built in to manipulate the document repositories and authorization systems of the external syndicated research providers as a way to eliminate the need to know the access rules locally.

COPYRIGHT COMPLIANCE

Typical enterprise search solutions, and their cousins the collaboration solutions, also struggle with copyright compliance when it comes to third party content. Internally produced content that these systems are founded on have no copyright issues to consider and as similar to the issues of third-party controlled subscription access rights discussed above, there are no basic constructs built into typical enterprise search and collaboration systems that can easily facilitate copyright compliance for external content.

But content licensed by external publishers to a restricted internal audience are always copyrighted, as is content harvested from external websites bearing copyright notices (as all do). Such content cannot be directly shared inside the company via a search result, a shared bookmark, or by the posting of a document to a collaboration site. The risk from well-meaning but clueless users manipulating enterprise search and collaboration solutions that are not informed on the copyright status of externally-acquired content is *\$50,000 per instance*.

Because each user who can access each copyrighted document is an instance for damage purposes, the damages can build up rapidly. For example, one case against a financial institution for widely and inappropriately sharing licensed syndicated research resulted in a \$450 million judgment against the company for copyright infringement that was upheld on appeal. Also, it should be pointed out that copyright infringement is an exception to the corporate shield, and a company's officers are personally liable for copyright infringement claims.

AGGREGATION OF LICENSED CONTENT ONTO THE ENTERPRISE NETWORK

For content licensed from external third-parties, it might seem that a possible solution to the aggregation problem would be to convince the publishers to permit the transfer of all the content to a repository on the company's network. Then it would look more like the normal on-network repositories the typical enterprise search engine is set up to handle.

There are several problems with this idea in practice. Very few of the content providers will permit it. The subject enterprise is the customer of the content provider and the content provider loses all control of access rights and usage metrics when the customer has the content onsite. Also, a typical content provider has 5,000 enterprise customers. The last thing most content providers want is the requirement to manage 5,000 dispersed repositories around the globe.

And if only a few syndicated research providers will not permit a full download of their content, or if even one crucial one will not, then the entire idea of an integrated search across all the sources fails. Also, in the rare case when all the providers contemplated today will permit a local repository of their content, the project is exposed to content providers changing their minds, or to a future desired sources not permitting it. Since it is very common for research firms to not permit local customer-controlled repositories, the risk of this situation developing as the research portal evolves is very high.

The process of aggregation from multiple external parties is operationally challenging to manage. Content providers are making constant changes to their publishing systems and document and metadata formats. IT departments are not “content factories” set up and staffed to monitor and make continuous changes to document repositories. Also, associated processes have to be created and managed for things like document updates, delete notifications, purges and legacy uploads when subscription options change.

Lastly, there is considerable financial liability for a company that wants to aggregate entire collections of syndicated research analyst content. Consider the dollar value of a syndicated research repository. A typical *SinglePoint* portal has one hundred thousand research reports worth a theoretical one-off price of maybe \$1,000 each. That is \$1 billion dollars in nominal value of the content. Imagine now that a security breach or disgruntled employee results in that repository being loaded from the client network to any of the many websites eager to open confidential material to the public. The financial liability of such an event is significant for a large company with deep pockets.

Working with Northern Light completely eliminates all of these problems and risks. Content providers do not object to Northern Light having brief access to their content because we are only one external party, not every one of their customers, and because we destroy our copy after we have finished indexing it. Northern Light insulates our client from the operational complexity of the aggregation process. And because the repository of content from all the syndicated sources is never downloaded en masse to the client, the client never has an onsite repository that puts the client at financial risk.

FEDERATED SEARCH VERSUS INTEGRATED SEARCH

One solution that might seem appropriate for getting around the problem of external content aggregation would be using a federated search solution instead of an integrated search solution like Northern Light. With federated search, the search engine sends the user query to the search engines native at each source, receives back a search result from each source (in, say, an XML package), munges the various search results together in some arbitrary fashion such as “first responder wins” and then presents them to the end user.

The attractions of federated search include the fact that federated searching obviates the need for a local repository on the enterprise network to index. And there are federated search connector vendors that will license a company the code required to put a syndicated source into a federated search solution. Also, Microsoft is supporting federated search in the ubiquitous SharePoint Search and has published a standard that all search engines at all content providers can adopt to join a SharePoint-based federated search application. Since the XML package of results provides a URL to invoke the document, users that are cookie'd for the content source should be able to click through to the document from the search result.

There are however, several critical flaws to federated search for strategic research portals. These include:

- Indexing strategies vary by content source. Some may provide proximity searching; others may not, so proximity searching is not practical in the portal research application. Some may provide case-sensitive search; others may not, so case-sensitive searching is not practical in the research portal solution. Some content sources may provide unstemmed as well as stemmed indexes (more on this later); others may not, so literal searching is not practical in the research portal solution. The list goes on and on.
- Supported search syntax varies by content source. Indexing strategy and search syntax are closely related ideas. The purpose of syntax is often to give access to specialized indexes. Without a common indexing strategy for the content in the research portal, it cannot be searched with the same syntax. In addition, some sources may support Boolean operators such as 'not' or complex Boolean expressions, while others may not. Or if they do, they may use different conventions for invoking a specific search function. Alerting can be very risky in federated search solutions because there is little visibility into how the various sources process queries. And a content sources supported syntax may change from time to time invalidating saved searches and alerts, like Google recently dropping the '+' operator that meant 'and' to everyone that has used Web search since Alta Vista introduced the '+' operator in 1994. This drives a lowest-common-denominator syntax with little more than simple 'and' and 'or' statements generally workable, if that.
- Accurate relevance ranking is impossible with federated search. Each of the search engines at the individual sources returns search results in ranked order using that search engine's relevance ranking method. However, the federated search results consolidator has no idea how any particular source's search engine arrived at the rank order. Each content provider can decide what elements and factors to weight and whether to rank them high or low for relevance ranking purposes. If there are 30 content providers in a federated search, there are 30 uncoordinated and hidden relevance ranking strategies in play.

Even worse, there is no rational basis for the process that munges the various search results from the different providers to determine how to interleave them for presentation to the user. How do you decide if hit number 3 from Source A is more or less relevant than hit number 30 from Source B or even hit number 300 from Source C? There is neither a statistical means nor information theory basis for making this determination.

So federated search engines use a variety of chance methods to do the munging including “first responder wins” (ironically rewarding the source that processes the query the least), random interleaving, source-quality based interleaving, and re-ranking based on the small bit of XML-returned search summary text for each hit. This last method is the current most popular with federated search connector vendors and also the one that produces the worst user experience since it destroys the one piece of reliable information in the search results from the various providers: the within-source rank.

- Taxonomies vary by content source. With federated search, taxonomies are the option of the content provider and no two will agree on structure, terms, extent, or any other facet of taxonomies. With federated search of disparate sources, the idea of a consistent taxonomy across all sources that reflects your research needs is unobtainable.
- Text analytics is impossible with federated search. In a similar manner to taxonomies, because the full-text of the documents is not available to the search engine in the research repository for indexing, there can be no text analytics applied to the documents searchable in the research portal. Text analytics operates by finding text strings and relationships between text strings that imply meaning. Without the full-text to operate on, this process cannot be carried out.

Northern Light has a completely different approach. We aggregate all the content by obtaining a full-text original copy of every document from every source. Then we index it with the *Northern Light Search Engine* using perfectly consistent indexing, taxonomy, and text analytic strategies uniformly across all sources. This produces an *integrated index* that can behave the same way across all sources in terms of search syntax and relevance ranking. Using Northern Light’s integrated approach:

- Indexing strategies are consistent across all sources. All sources have the same set of indexes such as stemmed, unstemmed, proximity, case-sensitive, and acronym.
- Search syntax is consistent across all sources. All sources can be searched using the same, advanced, flexible, and powerful search syntax. Alerts can also use this same consistent syntax and be depended on to work the same way for all sources. There is no risk that saved searches and alerts will be disabled for one or more sources by changes to supported syntax at a content source.
- Taxonomies can be applied across all sources. With Northern Light’s integrated search strategy, a client can use a relevant taxonomy applied to all sources in the research portal. Control of taxonomies resides entirely with the client, not with the content sources.
- Text analytics is enabled across all sources. Because the search engine has access to the full-text of the documents, powerful text analytics like Northern Light’s *MI Analyst™* are possible. Without text analytics, research portals are not using the most powerful tools available for speeding the time to insight.
- Relevance ranking is accurate across all sources. With integrated search, Northern Light can in fact tell if hit number 3 from source A is more or less relevant than hit number 30 from Source B or even hit number

300 from Source C. The relevance ranking formulas are applied to all documents in a consistent manner and every aspect of the document can contribute in the same way to the document's rank in the same way regardless of its source. There is no *munging* with integrated search, only *ranking*.

NORTHERN LIGHT SEARCH AND SEARCH-RELATED SERVICES

Northern Light has invested five hundred man-years in our *SinglePoint* application suite and the *Northern Light Search Engine* that is *SinglePoint's* technical foundation. For a client to come close to the level of functionality that we provide would require a sizable investment by the client in the portal UI, supporting systems, and content aggregation operations. Companies that have undertaken such projects, or at least estimated them, have reported budgets well into seven figures for the projects. By contrast, Northern Light *SinglePoint* costs one-tenth as much.

By contrast to the problems facing general purpose enterprise search solutions, the opportunities for the Northern Light Search Engine are highly defined by the known use case: strategic business research. Knowing the use case permits Northern Light to optimize the search technology and supporting enterprise search services in ways not available to general purpose enterprise search engines. Specifically, we can:

- Provide the standard range of functionality as all enterprise search solutions as identified on page 4 above.
 - Can index and search any common enterprise document type such as MS Office, email, PDF's, and HTML.
 - Can scale to reasonably high document counts typical of enterprise repositories.
 - Have acceptable relevance ranking.
 - Have extensive enterprise integration admin tools for doing things like identifying repositories, integrating with Active Directory for document security, etc.
 - Support industry standards for interoperability like OpenSearch, OpenAuth, Web Services, ODBC, federated search, etc.
- Provide content aggregation as a key service.
- Support third-party subscription access models.
- Enforce third-party copyright restrictions.
- Portal design that supports the needs of both casual research consumers and research power users.
- Extend the search indexes to useful additional types.

- Support extensive query syntax options.
- Optimize relevance ranking.
- Make maximum use of metadata.
- Make maximum use of taxonomies.
- Support really big queries.
- Provide optional customized thesaurus expansion.
- Make maximum use of text analytics.

Each of the above will be discussed in more detail in the chapters and sections that follow.

THIRD-PARTY EXTERNAL CONTENT AGGREGATION

PROVIDE CONTENT AGGREGATION AS A KEY SERVICE

As a core service, Northern Light aggregates business and technology research from external sources ingests it into the *SinglePoint* system to index and deliver documents to users. Northern Light's content aggregation capabilities span:

- Syndicated research publishers in every industry.
- News harvested both from the Web and from licensed industry news publishers.
- Scientific and academic journals.
- Life sciences conference presentations abstracts.
- Social media.
- Custom Web search (e.g., websites of a client's competitors)
- Government databases and regulatory information.
- Financial analyst reports and public financial data.
- Structured data from data vendors.
- Internal document repositories of primary research and market intelligence.
- Video content.
- Internal document repositories of primary research and market intelligence
- Custom content aggregation: any content in any format on any computer located anywhere in the world in the possession of any organization.

AGGREGATE AND SEARCH SYNDICATED RESEARCH ANALYST CONTENT

Syndicated research analysts often provide the key content set in strategic research portals. The names of the leading syndicated research firms vary by industry but in each case Northern Light works with them. For example, content partners of Northern Light include Forrester, IDC, Corporate Executive Board, Strategy Analytics, Celent, Reed Elsevier, Thomson Reuters, and Decision Resources. While the names change by industry, there are always twenty to fifty leaders in each industry (and some industries overlap) that provide highly valuable commentary, analysis, and perspective. Northern Light is currently technically integrated with 134 of these leading syndicated research providers to access, aggregate, index, search, and deliver research to clients. And we are adding more all the time as clients ask for new ones.

There are two points of technical integration with each content provider. First of all, Northern Light has to acquire the content for indexing. Secondly, the *SinglePoint* platform integrates with the document repositories and authorization systems of the content providers to deliver documents to users.

Thus, with two points of integration with each of our current syndicated research providers, there are 268 individual technical systems to create, maintain, and operate in order to make *SinglePoint* third-party content integration and delivery work for our clients. These points of technical integration are described below.

For the first point of technical integration with a syndicated research provider, Northern Light *creates a custom solution for acquiring each content set supplied by each third-party research provider* because, as discussed above, there is no commonality between the publishing systems, metadata conventions, document repositories, access rules, and content security systems of the various providers. These custom content aggregation solutions are worked out in close cooperation with each research provider. Common and easy industry solutions for aggregating content such as crawling the websites of the research providers or indexing RSS feeds of new report announcements do not work because the market research firms do not expose the full-text of their valuable, proprietary content to web crawlers and do not deliver the full-text of the documents in RSS feeds of report announcements.

Once Northern Light indexes a collection of content from any particular source, we can either place the content in our optional document repository or we can destroy our copy. *SinglePoint* is indifferent to the location of the document repository for any piece of content and can invoke it for authorized users regardless of where it resides, at a third party-content provider, on the client's corporate network, or in the document repository at Northern Light.

The importance of being able to aggregate, index, and search third-party syndicated research is best expressed by Northern Light's customers in the form of these statements made to a market research firm studying market research portals by Northern Light clients:

"It was a tremendous effort to get [the internal IT department] to see our vision, much less to implement it. But the biggest roadblock was just managing the vendors. Every vendor has a different set of legal

requirements. All those agreements were a tremendous obstacle to IT...It's amazing how convoluted the terms and conditions are. Northern Light maintains those agreements for us. They have established relations with all the research vendors and they have a track record. If I call up and want to add a new research vendor [to my SinglePoint implementation], Northern Light is usually already there. But if they weren't, they have an established process for making it happen. I really appreciate that."

"I wanted a vehicle where I could research multiple vendors at one time. It's a simple concept but a lot more complex to do...I went straight to IT. They could build it but managing the work behind the scenes just couldn't be maintained. It's not IT's job and we're not staffed to handle it. Northern Light is the only game in town. They do it all for us now including hosting."

"Where Northern Light excels is in their ability to get licensed access to IDC and Gartner. We're coming from a company that has a lot of experience with these vendors and we don't do it. Northern Light does it all. They update it. It's turnkey. The key is their relationship with the vendors. It would be prohibitive for us to manage all those relationships. We feel lucky that we 'fell' into Northern Light."

"[Northern Light] wrote a great piece on the *content factory* — that content is not an application as much as it is a process. Unlike, say, an accounting or email package that gets installed and keeps running, SinglePoint requires a daily refreshing of content to maintain its value. Every day, the content has to be aggregated, indexed, and unified. Material flows in from the outside, goes through a series of value-added processing steps to reach the finished product — utterly dependent on outside entities. It's true. Northern Light handles it all."

"Content integration is a service not a technology. It's not IT. IT will never want to do what SinglePoint does. IT can do the front end but no one but Northern Light does custom content integration. Every customer is custom integration. People don't always understand that."

SUPPORT THIRD-PARTY SUBSCRIPTION ACCESS MODELS

For most third-party content, Northern Light links users back to the website of the content provider to fetch the document when the document is requested by a user. This process is automatic, instant, and transparent to users. They just click on a link on a search result and the document appears in a browser window on their desktop. Behind the scenes, Northern Light logs the user into the content provider's website, navigates the website and content repository of the provider, uses the correct credentials to exercise the content provider's authorization

system, and places the requested content into a browser window. Users need not trouble themselves with the elaborate machinery and complex interacting processes occurring behind the scene to make it all work seamlessly.

There is no commonality between third-party research providers as to how their user authentication and document access systems operate. So just like in the content acquisition phase, each document fulfillment connection to a research provider has to be hand-tooled and maintained one research provider at a time. Northern Light does all the extensive development and monitoring necessary to make the access solutions for each content provider work. Delivering the documents from external third-party market research sources is a second point of technical integration with each third-party source.

As previously discussed, most third-party market research firms do not want repositories of their content scattered around on many other companies' networks outside of the control of the research firm. Most research providers have business policies against establishing such remote repositories of their content. Their willingness to cooperate with Northern Light is based in part on the indifference of the *SinglePoint* platform to the content repositories' locations. With our solution, the research provider can maintain the repository on the providers own network and we simply fulfill each document request from their repository.

And because we manipulate the research providers' own systems for document fulfillment using the credentials supplied by the research providers for this purpose, Northern Light cannot deliver content to unauthorized users because in the unlikely event *SinglePoint* tries to invoke a document for an unauthorized user, the invalid credentials will fail at the level of the content providers' authorization system.

Also, the transaction systems of the research providers see the content download event and the full texture of the content relationship between the provider and the consuming organization is maintained. All the anxieties on the part of the research providers are relieved.

In addition, the *SinglePoint* application is informed about the access rights of every users and every document in the repositories in the portal and advises the user on the user's access status even before the user tries to download a document from a research provider. This is necessary to providing the best user experience possible. We do not want a user to see relevant and interesting documents on search results; then try to download them only to discover that he or she does not have access to those documents. It is much better to give that information at the search result level.

ENFORCE THIRD-PARTY COPYRIGHT RESTRICTIONS

Northern Light *SinglePoint* has content access restrictions built in several ways. One has been described above in the discussion of subscription-rights enforcement. Since one way a company can violate copyright is to violate subscription access rights, *SinglePoint's* informed state on licensing rights simultaneously prevents copyright violation of licensed third-party content. This idea is extended in two ways in the *SinglePoint* application platform.

Firstly, there is the problem of web-harvested content that is freely available to users of the public internet. Sharing such content using links to the original website pages does not violate copyright as each new user the link is shared with simply visits the public website to view the content. *SinglePoint* shares all content via links at every point of contact (bookmarks, shared bookmarks, alerts, headline boxes, expert searches, etc.) so that such sharing is automatically copyright compliant.

Secondly, when it is desirable to have content that is reliably persistent, such as conference abstracts from scientific meetings, Northern Light can prepare a summary of the document and provide that to the client. Summaries are fair use under copyright law. These summaries can be rich with all the document citation metadata (title, source, authors, date), all the tags applied by Northern Light's *MI Analyst* text analytics solution, and a summary of the text on the original document.

BUSINESS NEWS

The second most important content set in many strategic research portals is business and technology news. News is key to tracking competitive intelligence, industry developments, technology trends, and customer intelligence.

NORTHERN LIGHT BUSINESS NEWS

Northern Light provides a business and technology news feed for *SinglePoint* clients that indexes 50,000 news stories per day from 5,000 hand-selected, high-quality sources. *Northern Light Business News* is especially rich in vertical trade news as trade journals have the timeliest, most relevant, and insightful coverage of industry events. For example, recently Northern Light's IT industry trade sites picked up the rumor on the new inexpensive iPhone days before that news found its way into the general business press.

Several Northern Light clients have switched entirely to *Northern Light Business News* from much more expensive licensed news sources, saving a great deal of money while improving the quality of their news feeds.

Using Northern Light's search and portal technologies, it is easy to customize and contextualize the news feeds for product groups, brand managers, key account teams, geographies, or for departments.

And if Northern Light is missing any sources that a client needs in the news feed, we will gladly add it for no charge so long as the content can be accessed from the Web for indexing. If the source requires a login, it falls into the licensed-news category.

LICENSED NEWS

Northern Light can integrate any licensed news feed into a *SinglePoint* portal using the same capabilities as we use for syndicated research publishers. We will technically integrate with the source to bring the content into the client's research portals.

SCIENTIFIC AND ACADEMIC JOURNALS

Northern Light currently indexes tens of millions of journal articles and abstracts from content sources like IEEE, Medline, PubMed Central, ProQuest, and EBSCO.

LIFE SCIENCES CONFERENCE PRESENTATION ABSTRACTS

Northern Light aggregates and indexes over 600,000 abstracts and posters from over 800 life sciences conferences around the world. Each abstract is manually reviewed to insure the metadata about authors and author affiliation is accurate and consistent.

SOCIAL MEDIA

Northern Light aggregates content from industry authority blogs, Twitter, Facebook, LinkedIn, and SlideShare. Social media is an important new source of expert commentary. For example, half of IT analysts mentions of companies are in social media and only half of the mentions are in their published syndicated research.

CUSTOM WEB SEARCH

Northern Light can crawl and index any part of the Web to make it searchable from a client's research portal. By far and away, the most popular choice would be the websites of a set of competitors. But other popular choices include Federal regulatory sites and thought leader websites.

GOVERNMENT DATABASES AND REGULATORY INFORMATION

Northern Light is expert at aggregating and indexing government databases, websites, document repositories, and news feeds into *SinglePoint* portals. Popular choices include SEC filings, clinical trials, patents, regulatory and compliance documents, and research databases.

FINANCIAL ANALYST REPORTS AND PUBLIC FINANCIAL DATA

Northern Light has a long and close relationship with Thomson Reuters that makes their repository of modestly embargoed financial analyst reports available for *SinglePoint* clients. These reports have valuable competitive intelligence and industry analysis, helped by the fact that the financial analyst authors have direct access to the CEO's of the companies they are writing about. The embargo period does not affect the utility of the report for strategic business analysis.

In addition, Northern Light can supply public financial data on companies from Thomson Reuters including a company snapshot, structured links to the company websites, stock price chart, current and historical financial results, officers and directors, and financial ratios.

STRUCTURED DATA FROM DATA VENDORS

Northern Light clients sometimes subscribe to market research and technology research in the form of data and not documents. Northern Light can access, search (e.g., metadata), format, and present such data in a *SinglePoint* portal.

VIDEO CONTENT

Video content is an increasingly popular content type. Northern Light can index the metadata of video repositories and download the video when the user clicks on it on a search result or can stream it to the user's default video player.

INTERNAL DOCUMENT REPOSITORIES OF PRIMARY RESEARCH AND MARKET INTELLIGENCE

Most *SinglePoint* portals have document collections presenting staff work on market research, competitive intelligence, product marketing analysis, and strategic plans. Internal documents come into *SinglePoint* portals using one of two general approaches:

- Uploaded manually using the included upload system. It takes about one minute of professional labor to upload a document including applying any required metadata. After working on a report for a few hours, days, or weeks, one minute to upload it at the end is not a burdensome requirement. The upload system can also be used for one-off purchased reports or commissioned research. Outside research contractors can be given access to the upload system so they can upload research to the *SinglePoint* portal directly.

- Automatically harvested from internal repositories. When document volumes are too large to make manual uploading a reasonable solution, Northern Light can aggregate the content using automated means in a fashion similar to how we automate the aggregation of syndicated research. For example, one client has 19 Microsoft SharePoint repositories with market intelligence that Northern Light aggregates using our SharePoint crawler solution.

CUSTOM CONTENT AGGREGATION

Northern Light can add any content in any format on any computer located anywhere in the world in the possession of any organization to a *SinglePoint* portal. So long as the client has legal rights to the material because the client licensed it, because it was freely available on the Web or in a government database, or because the client created it, Northern Light can aggregate it, index it, search it, and provide it to authorized users anywhere in the world.

We are especially good at finding solutions to difficult content aggregation problems such as those typical in dealing with Web content. While the Web has a gigantic amount of information on it, finding it, avoiding the extraneous and irrelevant material, capturing the necessary metadata, and extracting the valuable portions of the full-text can be a daunting task for the inexperienced client IT department or content publisher. Bring us your hardest content aggregation problems and we will solve them.

NORTHERN LIGHT SEARCH ENGINE

Up unto this point in the whitepaper we have discussed aspects of Northern Light's search related services including content aggregation and aspects of research portal optimization that are built into the *SinglePoint* portal application.

Now we turn to a discussion of the *Northern Light Search Engine* and related applications like *MI Analyst* text analytics and explain how they are optimized for market and technology research applications.

STEMMED INDEX

The *Northern Light Search Engine* provides the most extensive and useful set of search indexes in the search engine industry for business and technology research. Like most enterprise search engines, the *Northern Light Search Engine* provides a stemmed full-text index of the content to be searched. The words that are indexed in a stemmed index are stemmed to their roots and it is actually the roots that are indexed. The user query is also stemmed so that the user is actually searching on the stemmed versions of the user's query terms.

Stemming controls for singulars and plurals, word tense, gerunds, and participles. Stemming frees the user from having to think about alternative word forms and allows the user to search for "market trend" and find hits on "market trends" as well "marketing trend." The stemmed index is the one that is the default index serving most user queries and it does a pretty good job for most users much of the time.

Practices supporting correct stemming are:

- Language-specific stemming. Like most search engines, the *Northern Light Search Engine* detects the dominant language of the document being indexed and uses the correct stemming rules for that language. This means that when users specify their search language (which can be an option on the search form), the documents on the search result are more relevant.
- CJK character sets. Like most search engines, the *Northern Light Search Engine* supports indexing and search in the CJK (Chinese, Japanese, and Korean) character sets. These are particularly challenging as words are not separated by spaces in these languages and have to be inferred by context.

Many enterprise search applications stop with the stemmed index, while the *Northern Light Search Engine* uniquely goes significantly beyond stemmed indexing. Every document in any repository indexed by the *Northern Light Search Engine* has all of the indexes described below created in addition to the stemmed index.

UNSTEMMED INDEX AND LITERAL SEARCHING

The *Northern Light Search Engine* produces an unstemmed index for every document in the repository of research in a *SinglePoint* portal. The 'literal' operator directs that part of the query to the unstemmed index for resolution which can sometimes be helpful. Use of the unstemmed index interprets the query literally. For example:

patent and (literal: ("developing countries" or "developing country"))

The above query would find any variation of patent (patent, patented, patenting, patents) because that part of the query would use the stemmed index. But the second part of the query will use the unstemmed index and will be literally interpreted; only producing hits when the exact spelling "developing country" or "developing countries" are found. If the 'literal' operator were to be eliminated from the query above, hits would occur on documents having "developed country" because "developing" and "developed" would both stem to the same root "develop." Of course, documents discussing new products in *developed* countries are the exact opposite of what the user wanted.

PROXIMITY INDEX AND PROXIMITY SEARCHING

The *Northern Light Search Engine* records the location of every word and every concept in every document in the repository, creating the proximity index. There are actually two proximity indexes, one for stemmed words and one for unstemmed words. With proximity, the user can specify that the search terms be near each other with or without a specific order. Both the 'near' operator and the 'within' are supported syntax options. Note that the industry convention is, illogically to this writer, that 'near' considers word order while 'within' does not. (In common usage, the term 'near' would not indicate a preferred order.) Northern Light follows the search engine industry convention.

Use of the proximity operators directs that part of the query to the proximity index. It is a very good way to force close association with words without having to think of all the different ways things might be phrased. For example, a query of the form:

Microsoft near: 10 "virtual servers"

IBM within: 5 announcement

The first version finds documents that have Microsoft ten words or less away (in either direction) from the phrase "virtual servers."

The second query finds documents that have "IBM" preceding, by at least 5 words, of any form of the word "announcement" (announce, announced, announcing, announces, announcement).

PHRASE SEARCHING

Most search engines support phrase searching in some form, at least for short phrases. However, the proximity index also allows the *Northern Light Search Engine* to search for phrases of *any length* since a phrase is just a series of word-pairs one unit apart in the proximity index. For example, when a user enters the search term:

“bring your own device”

The *Northern Light Search Engine* translates the above user search into this technical version that is what actually gets processed:

((bring within: 1 your) within: 1 own) within: 1 device

This technique can be extended to phrases of any length.

CONCEPT INDEX AND CONCEPT SEARCHING

The *Northern Light Search Engine* indexes the concepts identified by our text analytics solution, *MI Analyst*, in all documents in the research repositories. For example, when “new product,” which might be expressed many ways, is identified in a research report or a news article by *MI Analyst*, the concept is indexed the same way as words or phrases are indexed. *MI Analyst* contains a taxonomy of tens of thousands of industry-specific concepts covering every aspect of business and technology. Once in the index, the concepts can be searched for as part of a search query. The desired for concept can be specified by use of the category name for the concept followed by the name of the concept. For example:

Business-issue: “strategic partnerships”

Global-risk: “hard landing of a developing economy”

NAME INDEX AND NAME SEARCHING

The *Northern Light Search Engine* produces a name index for every document in the repository. The name index only has the terms that have initial caps. The ‘name’ operator directs that part of the query to the names index. For example, a search for the retailer “Target” will be much more productive if it is phrased:

Name: Target

This query will return only those documents with a capital “T” as in “Target.” Documents that have a capital “T” in Target are likely to be about the retailer. Documents with only a lower case “target” will be skipped.

ACRONYM INDEX AND ACRONYM SEARCHING

The *Northern Light Search Engine* produces an acronym index for every document in the research repository that requires that all letters in the searched-for term be capitalized. The ‘acronym’ operator directs that part of the query to the acronym index. For example:

Acronym: HP

Searching for the acronym ‘HP’ avoids most hits on the abbreviation for horsepower which is often “hp” or “Hp.”

Or particularly useful in the IT industry is searching for the acronym ‘IT’ to avoid hits on forms of the word “it” (it, its, it’s) Most of the documents responsive to a simple query “IT” are not about information technology because the pronoun ‘it’ is so common. The problem of “it” is present even when “IT” combined with a relevant word to like “strategy,” as in this query:

“IT strategy”

The above query produces false positives on 85% of the returned search hits because the various forms of the word “it” in documents that also have some form of the word “strategy” right after the version of “it.” For example, the phrase “Burger King is considering its strategy....”

By contrast, consider this query:

(acronym: IT) within: 1 (strategy)

The above query produces virtually 100% precision, only returning hits about information technology strategy because both the “I” and the “t” will have to be capitalized as in the phrases “IT strategy” or “IT strategies.”

RELEVANCE RANKING OPTIMIZED FOR BUSINESS AND TECHNOLOGY RESEARCH

Unlike general purpose enterprise search engines that must serve many dissimilar use cases, the *Northern Light Search Engine* has the luxury of knowing that our users will be professionals doing business and technology research on large repositories of research reports and news stories. This allows Northern Light to optimize our relevance ranking for these purposes.

Research reports are different from webpages, emails, chats, or text-fields in SQL databases that other search engines from Google, Microsoft, or Oracle have to be able to search. Specifically, research reports have these unique characteristics:

- Large. Search results for very short documents like webpages, emails, and chat need not consider document length, structure, and text proximity in relevance ranking. Effective search of research reports which are usually many pages long (and can be a hundred pages long) does need to have these characteristics considered, as well as the proximity of the search terms to one another. Industry and technology news articles fall somewhere between an email and a research report in terms of size and they also benefit from considering things like document length.
- Rich with important metadata. Short user-generated documents like emails have little in the way of metadata that is useful. For example, an email about a key account might have a title like “tomorrow’s meeting” and not mention the key account in the title. As a result, general purpose enterprise search engines do not exploit metadata for relevance ranking as much as they might otherwise. By contrast, professionally-authored research reports and industry and technology news articles have rich metadata including meaningful titles, important authors, authoritative publishers, and context-setting publication dates. These metadata can be exploited for relevance ranking by a search engine, like Northern Light’s, that focuses on sources of these types.
- Rich With Terms-of-art and terms-of-context. Professionally-authored research documents and news articles are filled with terms-of-art (e.g., industry nomenclature for technologies) and terms-of-context (e.g., company terminology). Short user-generated documents like chat and email tend to avoid terms that are lengthy and don’t repeat them when other users in the threads have already used them. General purpose search engines often do not consider the uniqueness of terms and phrases in relevance ranking algorithms, or at least not as aggressively as would benefit the process.

The relevance ranking algorithm in the *Northern Light Search Engine* is explicitly designed to take advantage of the unique characteristics of research documents and industry and technology news. Northern Light’s relevance ranking balances these factors:

- Number of times the query terms are in the document. More is better and rewards more substantive documents typical of research repositories, but this factor has declining incremental impact as the number of mentions goes up to avoid overweighting really long documents.
- Number of times the query terms are in the document relative to the length of the document. This is a ‘density’ measure. It is generally better to have 25 hits on the query terms in a 5 page document than 10 hits on the query terms in a 200 page document.
- Word order and proximity of the query terms in the document. Users tend to search on naturally meaningful phrases. Therefore, even when the user did not force proximity by using phrase syntax or

proximity operators, word order and proximity can still be useful in determining relevance. For example, a user searches on

customer relationship management

In the above query, with no quotes to force a phrase, documents that have the phrase “customer relationship management” would be still be most likely preferred over documents that mention the search terms but with them separated from one another in the text, as in this example:

Paragraph 1: “*Customer* feedback is the key for input for effective market research about new products.”

Paragraph 5: “*Management* should strive to maintain an internal process that is consistent in researching and launching new products.”

Paragraph 25: Schedule reliability has a direct *relationship* on the ability of the company to launch new products.

When searching an email, a short call report, or a webpage on a SharePoint site, there may be very little text to start with so there is little benefit to general-purpose search engines to consider word order and proximity in relevance ranking. Proximity is a difficult factor to consider at runtime because it requires a proximity index (which most search solutions do not provide by default) and it is computationally burdensome to such a great degree that this factor is actively avoided by most search engines whenever possible.

Indeed, to our knowledge, Northern Light is the only search solution that considers word order and proximity heavily in relevance ranking. The *Northern Light Search Engine* always generates a proximity index, as well as generating other unique indexes useful for this purpose, and has a proprietary method of considering word order and proximity in relevance ranking that is not computationally burdensome, making no user- perceivable impact on query response time.

- Search terms in the document metadata. The *Northern Light Search Engine* recognizes the importance of metadata and boost documents with search-terms hits in the metadata like titles. This boost takes place even if the search terms as entered by the user did not explicitly include a title metadata search.
- Inverse document frequency of the search terms. The inverse document frequency (IDF) of the search terms measures how rare each search term is in the database as a whole and rewards documents that are relatively richer in the relatively rarer search terms. The rare words are better discriminators for relevance ranking purposes. So, for example, consider this query:

Microsoft and servers

“Microsoft” will be the most common word in a database of IT analyst research reports and knowing a document has Microsoft in very often and everywhere will contribute little meaningful intelligence to relevance ranking since many, many other documents will too. By contrast, “servers” will be rarer and using that term more heavily will be better able to get the best documents to the top of the list.

IDF is a well-known technique for relevance ranking and the literature abounds in discussions of it. Other enterprise search engines use the IDF as a relevance ranking factor, but short documents like emails, webpages, or chat logs have less need for it. Northern Light’s impression is that we “dial up” the impact of IDF by far more than many other enterprise search engines, with the IDF weighting being a good example therefore of our tuning the *Northern Light Search Engine* for our use case.

- Date. All things being equal, the *Northern Light Search Engine* ranks more recent documents higher than older documents. In the default setting, this factor is weak and may have little impact because date sort and date-range select are both supported and usually preferable options for using dates. However, the impact of date boost can be turned up for clients that want to bias the results to more recent documents.

Taken together, the above factors are the Northern Light “secret sauce” creating search results for research repositories of extraordinary quality. Clients that have compared Northern Light’s search results with legacy systems or commercially available enterprise search engines have, over and over, reported easily observable and significantly better quality of Northern Light’s results. Below are some direct quotes from Northern Light clients about the stellar search results they have experience from the *Northern Light Search Engine*:

"Northern Light is a class act. A niche’ company with **deep knowledge of search, relevance**, and aggregation as it relates to news, the analyst community, and industry white papers from several industries. Laser focus on research."

"Excellent company, **with cutting edge search** and integration capabilities. Also, the customer service is above excellent."

"Our professional secondary researchers were very skeptical. They didn’t believe that consultants would use it or that the search capability would be as good or better than the native searches on each vendor site. In fact, there is a very straightforward user interface — very intuitive. We’ve had to do little or no training whatsoever. And **the search is first rate**. They’re won over."

"Northern Light has **a great pedigree for search and has a first-class capability** — for sure. "

“I was impressed by [Northern Light’s] knowledge of information. It’s really deep and practical — not just about technology, but about how information is used. Most search companies know search. And most content companies know content. Very few people understand the way that information is used. [Northern Light is right there at the intersection of content and technology.](#)”

The last quote above was from the Director of Knowledge Management at a Northern Light client and that individual currently has the same title at one of the largest IT services consulting professional partnerships in the world. The quote is included because it illustrates the key reason Northern Light’s relevance ranking is better than other enterprise search engines for business, industry, and technology research.

As smart as our engineers are (and they are very smart), Northern Light does not employ engineers who are more knowledgeable about search algorithms than the engineers at firms like Microsoft or Google. Rather, Northern Light understands the use-case of research better because we have, as the very first client above observed, a laser-focus on that use-case. And because we own the search engine, we can evolve the programming code to better match our clients’ needs for the use we know for sure they are going to put our search engine to.

CUSTOMIZABLE RELEVANCE RANKING

While our defaults are usually the best settings for a client, relevance ranking in the *Northern Light Search Engine* can be tuned to an individual client’s needs. There are some standard items that can be adjusted with a simple parameter setting, these being:

- **Title Boost.** Turns up the role of search term hits in the title in relevance ranking. Titles have profound impact on users’ perceptions of relevance and the easiest way to impress casual users quickly is to turn up title boost. Power users might prefer the default setting as it will most times produce more relevant documents overall, but you have to look inside the documents to tell that.
- **Date Boost.** Turns up the role of recency in determining relevance ranking. As mentioned above in the section on relevance ranking factors, in the default setting, this factor is weak and may have little impact because date sort and date-range select are both supported and usually preferable options for using dates. However, the impact of date boost can be turned up for clients that want to bias the results to more recent documents.

In addition to the above, a client can define any metadata that the client feels should have additional relevance considerations. For example:

- Document Collection. This can be useful for providing “best bets.” For example, internal market research might be higher than syndicated research.
- Document Source. If there are authoritative sources, they can be biased toward the top of the search results. For example, *FiercePharma* might be preferred over other life sciences news sites or *Reuters* over other business news sources
- Tags. For example, Portal admins can tag key documents as they are uploaded and the tags can be weighted heavily by the relevance formula. Different tags can receive different weight adjustments.
- Any metadata. Any metadata in a document repository can be used to weight documents on search results. PowerPoint can be preferred over HTML documents. Documents with project numbers might be preferred over documents without them. Documents might be preferred that contain a concept like “patent litigation “ over documents that don’t independent of whether the search terms refer to patents or to litigation. Departments from the CEO might be preferred over those he or she didn’t write. The options are endless on only limited by the client’s need and imagination.

EXTENSIVE QUERY SYNTAX OPTIONS

Power users whose jobs are research often understand and can take advantage of advanced search syntax. Also, research portal administrators can use the rich syntax options to create powerful and tuned search queries that can be used as user-subscribable expert searches, dashboard queries, and headline box queries. In this manner, the benefits of the advanced query syntax options can be extended to the entire organization and not reserved to just the power users.

The *Northern Light Search Engine* supports the richest array of search syntax available in the search engine industry, as described in the sections that follow.

COMPLEX BOOLEAN EXPRESSIONS

The *Northern Light Search Engine* supports all Boolean operators: AND, OR, NOT. Compound and nested expressions can be formed to make complex Boolean expressions for highly-refined searches. For example:

(IBM or Oracle or Sun or Dell) and (SaaS or cloud) not (consumer or “small business”)

The above query will find documents that mention any of the companies listed along with SaaS or Cloud computing, with the added restriction that the reference not be to a consumer or small business market.

WILDCARDS

The *Northern Light Search Engine* supports wildcards, which can be helpful in certain situations where stemming to roots does not produce the desired flexibility for finding alternative word forms. In many cases, this occurs when slang terms have developed to substitute for longer words and the slang terms are not in standard root maps. For example, “pharma” does not have “pharmaceutical” as a root so the use of a wildcard might be handier than typing out the alternative. Or chemical or scientific names often have similar beginnings but varied endings. Queries using wildcards are useful in these circumstances, for example:

(“customer relationship management” or CRM) and pharma*

Rather than ...

(“customer relationship management” or CRM) and (pharma or pharmaceuticals)

Or ...

deoxy*

Rather than ...

deoxyribonucleic or deoxyribose

METADATA INDEXING AND SEARCH

The *Northern Light Search Engine* creates explicit indexes for all metadata of interest. Typical metadata fields where the ability to search is often useful include:

- Title
- Source
- Author
- Author affiliation
- Date published

- User applied-tag
- Project ID
- Document type (e.g., Word, PowerPoint, PDF)
- Report type (e.g., Quarterly Market Share Update)

There is no limit to the number of metadata tags that can be applied. Placing metadata into searchable indexes can be very powerful. Using the relevant metadata index can be used to search, filter, group, and navigate search results. Search forms can be designed to expose the metadata for users, even casual users, to utilize to filter search results, or syntax can be used by power users in the 'search for' field, or *MI Analyst* categories can be used to provide faceted navigation of the search results based on document metadata.

COMBINED SYNTAX

The *Northern Light Search Engine* supports the ability of users to combine any and all of the index and syntax options into a single search term. Using the general form of complex Boolean expressions the various syntax options can be used to create search queries of great power to focus results on the most relevant documents.

For example:

(price near: 5 (cut or slash or reduce or lower or reduce)) near: 20 Oracle

This query cleverly has a nested Boolean expression and two proximity operations! The query provides for all the common ways a document might use word meaning 'price cut' (for example "price cut," "cut price," "lowered the prices," "slashed prices," "reduced the product's prices," etc.). Once the conditions for a price cut have to be met by relating the term 'price' to any of a series of specified words, then the location of that condition has to be near the term Oracle in the document.

Another example:

(title: Oracle) and ("database servers" near: 10 announcement)

In this query, the user is combining a metadata search in a compound Boolean expression that has a phrase linked by a proximity operator to another search term. This query will find any document that is *about* Oracle and which contains a discussion of database server announcements.

One last example of combining syntax and search options:

(name: Target) and (source: (“retailing today” or “inside retailing”))

This query will return documents highly likely to be about the retail giant Target and that are from the news sources *Retailing Today* or *Inside Retailing*. Restricting the query to the names index reduces the likelihood that spurious hits will be produced by articles about companies *targeting* retail strategies. The query combines use of a special index (the name index) with a metadata index. Note that there are no search terms in this query against the full text, which is often helpful if the goal is to, say, monitor *all* the news about a key account from authoritative publications that follows the key account closely.

REALLY LARGE QUERIES

Many search engines limit the query length to speed up processing. Google for example, ignores all words in a query beyond the first 32, including operators such as ‘or’ or ‘and.’ For strategic research, limiting query length can limit the effectiveness of search results. For example, if there is a desire to monitor a list of 20 competitors, there would be a minimum of 39 words (each name with ‘or’ between them).

The *Northern Light Search Engine* has no hard and fast limit on the number of characters in search query. Really long queries can be created (and saved for reuse) for those cases in which monitoring a long list of topics or companies is useful. For example, this query would be impossible (or would be secretly truncated by the search engine), but it works very well on the *Northern Light Search Engine*:

(new near:10 product) and (IBM or Oracle or CSC or Dell or Microsoft or Amazon or Intel or Fujitsu or Accenture or Capgemini or Hitachi or Ericsson or (acronym: NTT) or (acronym: NEC) or “BT Global Services” or “T-Systems” or Lockheed or SAIC or Atos or Huawei or Siemens or “Digital China” or Tata or TCS or Logica or “General Dynamics” or EMC or Alcatel or Cisco or (acronym: SAP) or (name: Sun)

The above query has 73 words and 397 characters including spaces. This query could not be successfully executed by the Google enterprise search appliance or any enterprise search engine that restricts query length severely. By contrast, the *Northern Light Search Engine* easily executes it.

It should be noted that Internet standards imposed unilaterally by Microsoft do limit effective query size. MS IE9 has a limit on URL length of 2083 characters. The *Northern Light Search Engine* sends the search terms to our query servers as a URL, so the limitation applies to us by default. There is some overhead (about 50 characters worth) with any search sent to the query server, so the actual number of characters in the query should be limited to 2,000 to always be safe. At an average of 5.5 characters per word, queries of 350 words are practical, which is over ten times larger than Google and other query-limiting enterprise search engines permit

THESAURUS EXPANSION

The *Northern Light Search Engine* can expand user queries to equivalent terms. With a thesaurus expansion, a user's query is expanded automatically to include equivalent terms. For example:

BYOD can be expanded to: (BYOD or "bring your own device")

Laptop can be expanded to: (laptop or notebook)

"Consumer products division" can be expanded to: "Consumer Products Division" or CPD

Any number of terms can be equated to one another. Any equated term will be expanded to all the other equated terms.

This feature is customized for each *SinglePoint* client because the appropriate expansions are often client-specific terms. Northern Light does not believe that weak associations provide a good user experience for thesaurus expansions, and recommends that clients only select highly equivalent terms for the expansion. Weak associations confuse users when the terms they searched for are not in the documents and weak associations produce false positives as well.

For example, these might be weak associations:

"Computer" expanded to also search for "servers" - many computers are not servers

"Phase III" expanded to "clinical trial" - many clinical trials are not phase III, but are phase I, II, or IV

"Federal budget" expanded to also search for "Federal spending" - the overlap of news articles using both these terms is only 15% and they address different issues (most of the articles about "Federal spending" are on specific programs while most of the articles about the "Federal budget" address aggregate government spending across all programs)

In every case, Northern Light works with the client to produce a search term thesaurus that is specific to the client's organization and industry. Also, *MI Analyst* concepts can be used for thesaurus expansion because the synonyms are explicitly used for tagging documents with concepts in *MI Analyst*. These adhere to the "equivalent terms" principle.

FACETED NAVIGATION

One of the most useful features of any search solution is faceted navigation through the search results. A typical search on a robust document repository returns hundreds or thousands of times more hits on the search results than the user will have time to examine. Faceted navigation exposes characteristics of the documents on the search result and allows users to navigate down through the search results based on these characteristics.

The power of faceted navigation is that the user does not have to correctly anticipate the options for refinement while writing the search query. Rather, the user can examine the options for refinement by observing the facets that are part of the search results and then refine the search results after seeing what interesting possibilities are present given the documents that were responsive to the initial query.

The *Northern Light Search Engine* provides faceted navigation through search results based on *MI Analyst* text analytics concepts. A fuller description of *MI Analyst* occurs later in this paper but for now we can observe that there are dozens of concept categories that clients can choose from, many of which are specific for the client's industry, and some of which are general in nature. Also, Northern Light typically customizes *MI Analyst* concept categories for each client. Brands, product lines, and hot topics are typical customizations.

In the screen shot below, the user has the option to navigate through the search results based on facets "business issues," "company" (mostly covered by the expanded window), "venture funded company," "IT technologies," or "government agency."

SinglePoint Best Practices Demo Portal

Dashboard | My SinglePoint | My Research | My Searches / Bookmarks | My Network / Groups | MI Insight Radars | Wiki | My Profile | Contact | Help | Logout

Sample Research Content | Business News | White Papers | IT Analysts Social Web

You are logged on as David Seuss (HP)

Search Results: Sample Research Content Collection

Analyze Your Search | Search for: tablet computers | Search | Refine This Search | Save this Search as an Alert

Business Issues

- Disruptive Products and Pricing (91)
- New Products (66)
- Market Segmentation (28)
- Service and Support (25)
- Professional Services (20)
- Best-In-Class Products (18)
- Product Differentiation (13)
- Target Market (12)
- Sales Cycle (7)
- Niche Market Focus (6)

Venture Funded Companies

- Twitter (7)
- Pandora (6)
- Fortinet (5)
- Groupon (5)
- Dropbox (4)
- Siri (4)

IT Technologies

- Tablet Computi...
- Smartphones (...)
- Mobile Computi...
- Cloud Computi...
- Social Networki...
- Android Operati...

Government Agencies

Search results for "tablet computers" (91 results, last 12 months). Sort by Relevance / Date / Mentions.

1. Disruptive Products and Pricing: --7, Forrester Research: Licensed Content, 08/25/2011

2. New Zealand Telecommunications 2011-2015 Forecast and Analysis: Irresistible Market Forces Meet Immovable Politics (pdf) | Actions | Mentions of "Disruptive Products and Pricing": --5, IDC Market Research: Licensed Content, 10/13/2011

3. Harmonic - Digital Media Infrastructure | Actions | Mentions of "Disruptive Products and Pricing": --3, Current Analysis: Licensed Content, 05/30/2012

Logos: FORRESTER, IDC, Current Analysis

INTEROPERABILITY WITH MICROSOFT SHAREPOINT AND OTHER ENTERPRISE SYSTEMS

Northern Light holds interoperability as a core concept in our technology. The *Northern Light Search Engine* supports all industry standards for interoperability including:

- Microsoft SharePoint Web Parts collection for embedding of search facilities and research content from the *SinglePoint* portal throughout the enterprise's Microsoft SharePoint installations.
- Web Services API for supporting the embedding of search facilities and research content from the *SinglePoint* portal throughout the enterprise in non-SharePoint application environments.
- Microsoft SharePoint Crawler for harvesting content for indexing from internal market and technology research SharePoint farms into *SinglePoint*.
- Support of Microsoft SharePoint Federated Search so that *SinglePoint* can embed content from the research portal into SharePoint Search federated search results.
- Support of SSO, including industry standard SAML, and integration with Active Directory to authenticate users for *SinglePoint* portals and personalize the experience for each user on the fly.
- Support of industry standards like OpenSearch, OAuth, ODBC, and WebTabs (Salesforce.com) for non-SharePoint Search applications. (Even though in fairness, SharePoint supports OpenSearch as well.)

Northern Light is pioneering the automated extraction and discovery of meaning from large repositories of research content.

Our meaning extraction solution is called *MI Analyst* and it is applied to every document and repository in *SinglePoint* portals. It provides these benefits:

- Automatic tagging of all documents from taxonomy of tens of thousands of concepts.
- Enhanced search result display giving the users more information on each document.
- Ability to navigate large search results sets using concepts.
- Automatic identification of relationships between concepts in the search result, i.e., scenarios of meaning.

And as a note to the reader, *MI Analyst* is pronounced “my analyst.” The “MI” stands for “market intelligence.”

THE PROBLEM

Let’s consider an example in the use of search applications by professional users. Imagine a company is considering a move into the Internet telephony business. For competitive analysis purposes, a member of the market intelligence staff at the company decides to analyze the strategy of Cisco Systems in “voice over Internet protocol” (VOIP). Assume there is a content repository available to the user of a few hundred thousand market research reports from dozens of authoritative information technology market analysis firms like Forrester, IDC, Current Analysis, Strategy Analytics, and others.

When this user searches for “Cisco and VOIP” in the content repository of IT analyst reports using an enterprise search engine of the current generation from the regular providers, the search result will list thousands of reports. Having produced this lengthy search result, the search engine then washes its hands of the situation, metaphorically dumping the pile of documents on the user’s desk and saying, “Cya!” as the search engine bolts out the office door. The user is left to sort through the pile to find at least a few documents the user thinks might be interesting to read.

Because one cannot know what one did not find, there is no objective way for the user to assess whether the documents that he or she actually took the time to read comprehensively represent the body of knowledge contained in the thousands of returned documents on the search result. What the user is actually doing is desperately hoping that the few documents he or she selects to read contain all the important findings, analysis, and perspective available on the topic.

INTRODUCING MEANING EXTRACTION

So how might search work better? One way is by applying *meaning extraction*, an emerging technology that identifies concepts contained within documents and document repositories, and surfaces combinations of these concepts that imply meaning in the context of the business, professional, or technical purpose of the search process. Today, meaning extraction is beginning to be applied by companies to search electronic document repositories and various online resources to dramatically improve and accelerate a researcher's ability to gain insight into a topic and answer specific research questions.

Meaning extraction works as follows:

Extract references to important concepts from every document in the research repository, particularly concepts that imply meaning for the business or professional purpose of the search.

- Record the location of each concept in each document in the research repository.
- Identify patterns of proximity-related concepts that imply meaning to a knowledgeable practitioner.
- Analyze the documents responsive to a search query to identify those patterns and highlight them to the user.

MEANING-LOADED ENTITIES

Entity extraction itself has been around the text analytics world for almost two decades. Traditionally, the entities being extracted are proper nouns, specifically: people, places, and organizations. For example, these are the default entities tagged in SharePoint Search from Microsoft. Using such a text analytics solution, the search engine can tell you that Cisco is in a news story, or in ten thousand news stories in the news article repository.

The first useful extension to entity extraction contained in Northern Light *MI Analyst* is to include a relevant taxonomy of *context-specific entities* that go beyond the proper nouns used in traditional text analytics. In an information technology setting, these context-specific entities might be technologies, for example VOIP, cloud computing, or software as a service. In different settings the relevant set of context-specific entities would be different. For example, in a pharmaceutical setting one might include context-specific entities like diabetes, Lipitor, or monoclonal antibodies.

The second and more profound extension to the idea of traditional entities that Northern Light made with *MI Analyst* is the idea of *meaning-loaded entities*. Meaning-loaded entities have depth and purpose-driven relevance. Meaning-loaded entities are events, conditions, situations, outcomes, actions, relationships, and trends that imply significance for the professional purpose of the search. For example, in a market intelligence search application,

meaning-loaded entities might be *price cut*, *change in market share*, or *strategic partnership*. In a pharmaceutical research setting, a meaning-loaded entity might be *clinical trial* or *generic drug approved*.

MI Analyst indexes and records the locations in all documents to all references to traditional entities (e.g., Cisco), context-specific entities (VOIP), and meaning-loaded entities (e.g., strategic partnership). Just for shorthand, let's refer to these three entity types collectively as *concepts*.

An effective meaning extraction application requires tens of thousands of concepts to facilitate the meaning discovery search process. As the concepts are identified they are organized into taxonomy - a *meaning taxonomy*, if you will. (Practically speaking, the meaning taxonomy usually precedes the concept identification.)

The meaning taxonomy is designed using a hierarchy that is specifically relevant to the context. For example, VOIP would be placed into the IT Technologies node of the meaning taxonomy while strategic partnership would be placed in the Corporate Strategy node. For the pharmaceutical example, diabetes is placed in the Diseases node, Lipitor in the Drugs node, and monoclonal antibodies in the Proteins node.

MI Analyst provides meaning taxonomies with tens of thousands of concepts for all the industries Northern Light serves with *SinglePoint* portals. If we acquire a new client in a new industry, we extend out meaning taxonomies to incorporate the needs of the new client. Also, custom concept branches are common in the taxonomy reflecting clients' brand names, product categories, and hot topics.

CONCEPT IDENTIFICATION, NAVIGATION, AND SEARCH

Meaning extraction exposes the concepts found in documents responsive to a search query to the user at both the document level and the search results level.

At the document level, *MI Analyst* presents the concepts found in a document as an enhancement beyond the all-to-brief document summary of the style supplied by traditional search engines. This assists a user in gaining an at-a-glance understanding of what is really in the document so the user can make a more informed decision about whether this report or journal article should be downloaded and read. This facility helps a user find those reports and articles that are most likely to be of the most value, which is crucial considering that only a few documents from a long list of search results are actually going to be read.

At the search result level, identifying the concepts found in all the documents on the search result represents an overview of the knowledge that is contained in those documents. Such a summary overview provides an opportunity for knowledge discovery that can surprise the user with insights otherwise unavailable.

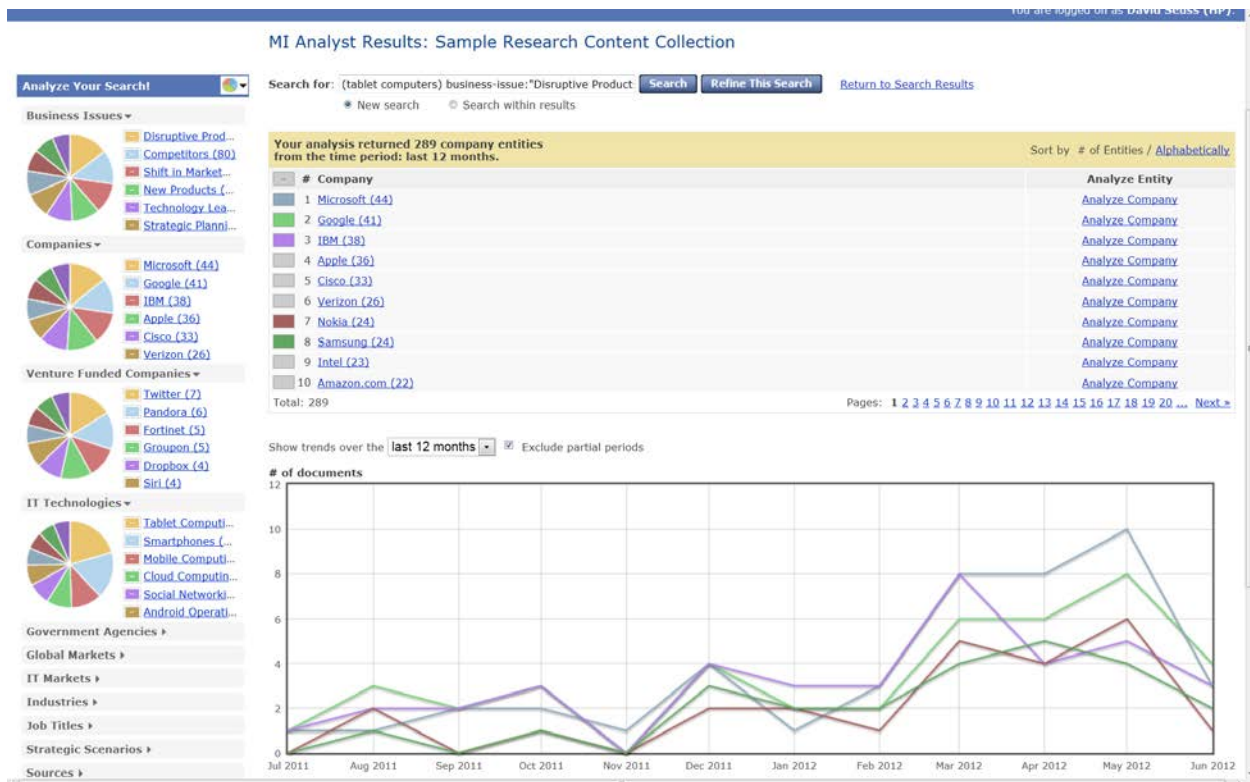
Also, it is possible at the search result level to use the concepts to provide faceted navigation of the search results that is discussed in the previous chapter on Northern Light search technology. Faceted navigation facilitates drilling down through successive layers of content by, in effect, linking concepts by clicking on them in the display

of the taxonomy that appears beside every search results. For example, one could start with a general query, look at the business strategy concepts returned, identify a relevant issue, and drill down into that one to find documents about that company using that strategy. In the example below (previously shown in the chapter) on Northern Light Search Technology, the user is drilling down through IT analyst reports that identified “disruptive products and pricing” as a strategy concept:

The screenshot displays the 'SinglePoint Best Practices Demo Portal' interface. The search results are for 'tablet computers'. A taxonomy of business issues is shown on the left, with 'Disruptive Products and Pricing' selected, showing 91 results. Other categories include New Products (66), Market Segmentation (28), Service and Support (25), Professional Services (20), Best-In-Class Products (18), Product Differentiation (13), Target Market (12), Sales Cycle (7), and Niche Market Focus (6). Below this, there are sections for 'Venture Funded Companies' and 'IT Technologies'. The search results list several documents, including one from Forrester Research titled 'Disruptive Products and Pricing' and another from IDC titled 'New Zealand Telecommunications 2011-2015 Forecast and Analysis: Irresistible Market Forces Meet Immovable Politics'. The Forrester document discusses the proliferation of tablets and smartphones in the workplace. The IDC document discusses the New Zealand market for mobile devices, mentioning tablets and smartphones.

MI Analyst also displays an automated trend analysis of the concepts found in the documents that come back on a search result. This helps users identify changes in coverage that might indicate an event of interest. Users can click on a date node in the graph and instantly drill down into a subset of documents from that date range.

In the example below (continuing the drill down from the page displayed above), coverage of IT technology tablet computers mentioning the companies Microsoft, Google, and interestingly, IBM, all jump when Microsoft announces its plans for Surface as a potentially disruptive product.



Also previously identified in the chapter on Northern Light search technology, *MI Analyst* also provides concept searching. All *MI Analyst* concepts are indexed, and once in the index, they can be used to search for documents. For example, as part of a search query one could search on a corporate strategy concept with this syntax:

corporate-strategy: strategic partnerships

AUTOMATIC IDENTIFICATION OF SCENARIOS OF MEANING

After the concepts are identified and organized into the meaning taxonomy, the next step is to interpret combinations of concepts as potentially significant. A human practitioner in the relevant knowledge domain specifies patterns of concepts that when found in proximate relationship to one another imply meaning to the professional researchers using the search engine. Let's call the relationships among concepts that fit the specified patterns *scenarios*. Because the scenarios can be specified at the taxonomy-node level, a pattern efficiently expressed in simple terms by the human expert can expand at runtime into a search for many, many specific relationships between individual concepts.

MI Analyst analyzes all the documents on a given search result and identifies the scenarios, flagging those found for the user to review. This automated analysis to identify scenarios represents the power of the machine to leverage human intellect. For example, in one deployed pharmaceutical application, *MI Analyst* looks for 1.9 trillion

potential scenarios on every search against a repository of 25 million journal articles, returning those scenarios found in a given user query in less than ten seconds.

The efficiency of using meaning taxonomies to create scenarios is illustrated by the fact that the text specifying the above scenario pattern consists of only 38 words that instruct *MI Analyst* to look for relationships between all of the entries in the taxonomy nodes for drugs, diseases, cells, cell receptors, medical devices, proteins, enzymes, genes, and therapeutic strategies. Then for each document in the 25 million journal article repository that is returned on a search result, the analytical process in the meaning extraction step examines all combinations of concepts in all specified taxonomy nodes to find those concepts that are in proximate relationship to one another according to the pattern of the scenario specification.

The relationships found by the meaning-extraction enabled search engine should be considered *scenarios*, rather than *conclusions* or *findings*, since it is impossible at present to automatically determine if the identified relationships are obvious, spurious, or significant. That question is left for the human intellect of the user to ponder. *MI Analyst* can only determine that the scenarios are *present*, and it can measure the number of documents each scenario is found in, which is the single most helpful indicator of *weight*.

Once the scenarios have been identified from all the tens of thousands of reports or journal articles on a search result, they are presented to the user to consider. In the example we started with, a search for Cisco using IT analyst research reports using this query:

Cisco and VOIP

These actual search results were produced by *MI Analyst*:

The screenshot shows a search results page for the query "Cisco and VOIP". The page is titled "Strategic Scenarios mentioning company: Cisco". It features a navigation bar at the top with tabs for "Dashboard", "My SinglePoint", "My Research", "My Searches / Bookmarks", "My Network / Groups", "Thomson Reuters Company Radar", "Events Radar", and "Wiki". Below the navigation bar, there is a search bar and a list of search results. The results are organized into three main categories: "Company Technology Focus", "Corporate Strategies", and "Product Marketing Strategies". Each category contains a list of search results with their respective counts. For example, under "Company Technology Focus", there are results for "Cisco is working on the technology Cloud Computing (207)", "Cisco is working on the technology Voice over Internet Protocol (VoIP) (172)", "Cisco is working on the technology Routing & Switching (100)", "Cisco is working on the technology Voice Telecommunications (91)", "Cisco is working on the technology Private Branch Exchange (PBX) (88)", "Cisco is working on the technology Mobile Computing (70)", "Cisco is working on the technology Network Services and Infrastructure (70)", and "Cisco is working on the technology Tablet Computing (69)".

Dashboard | My SinglePoint | My Research | My Searches / Bookmarks | My Network / Groups | Thomson Reuters Company Radar | Events Radar | Wiki | My Profile | Contact | Help | Logout

Business Research · Business News · White Papers · IT Analysts Social Web

You are logged on as David Seuss (HP)

Strategic Scenarios mentioning company: Cisco

[Company Technology Focus](#) | [Corporate Strategies](#) | [Cost Cutting](#) | [Financial Issues](#) | [Global Risks](#) | [Legal Issues](#) | [Macro-Economic Trends](#) | [Marketing Issues](#) | [Patent Strategy](#) | [Problem Situations](#) | [Product Marketing Strategies](#) | [Strategic Planning and Execution](#)

Concepts mentioned in documents containing **Cisco** and your query **((Cisco and VOIP) company:"Cisco") company:"Cisco"**

Company Technology Focus

- [Cisco is working on the technology Cloud Computing \(207\)](#)
- [Cisco is working on the technology Voice over Internet Protocol \(VoIP\) \(172\)](#)
- [Cisco is working on the technology Routing & Switching \(100\)](#)
- [Cisco is working on the technology Voice Telecommunications \(91\)](#)
- [Cisco is working on the technology Private Branch Exchange \(PBX\) \(88\)](#)
- [Cisco is working on the technology Mobile Computing \(70\)](#)
- [Cisco is working on the technology Network Services and Infrastructure \(70\)](#)
- [Cisco is working on the technology Tablet Computing \(69\)](#)

Corporate Strategies

- [Cisco is using a corporate strategy of Strategic Partnerships \(96\)](#)
- [Cisco is using a corporate strategy of Channel Partnerships \(60\)](#)
- [Cisco is using a corporate strategy of Acquisitions \(54\)](#)

Product Marketing Strategies

- [Cisco is using a product marketing strategy of Professional Services \(65\)](#)
- [Cisco is using a product marketing strategy of New Products \(32\)](#)
- [Cisco is using a product marketing strategy of Service and Support \(31\)](#)
- [Cisco is using a product marketing strategy of Market Segmentation \(30\)](#)

- [Cisco is working on the technology Smartphones \(65\)](#)
- [Cisco is working on the technology Virtualization \(64\)](#)
- [Cisco is working on the technology Web Conferencing \(60\)](#)
- [Cisco is working on the technology Video Conferencing \(48\)](#)
- [Cisco is working on the technology Collaboration and Groupware \(46\)](#)
- [Cisco is working on the technology Session Initiation Protocol \(SIP\) \(45\)](#)
- [Cisco is working on the technology Software as a Service \(SaaS\) \(40\)](#)

[Return to Top](#)

[Return to Top](#)

Summarizing the most numerous scenarios in the strategies section (and making more legible since the screenshot is hard to read):

- Cisco is using a corporate strategy of Strategic Partnerships (96 reports mention)
- Cisco is using a product marketing strategy of Professional Services (65 reports mention)
- Cisco is using a corporate strategy of Channel Partnerships (60 reports mention)
- Cisco is using a corporate strategy of Acquisitions (54 reports mention)

It's immediately obvious that these scenarios are no ordinary search results; Cisco's strategy in the VOIP market jumps right off the page without reading a single document. *MI Analyst* is suggesting that Cisco is using a combination of partnerships/acquisitions, presumably to penetrate the market quickly, and high levels of professional services and support as its lead product marketing campaign. Each of the search results listed above is linked to a list of reports that discuss that scenario, sorted by the number of times the scenario is in the report so users can rapidly drill into the documents that best elaborate on Cisco's strategy. So, for example, the user could drill down into the subset of reports that present target markets and market segments to learn what those target market segments are.

Also, consider that had the user in the example above used a standard enterprise search engine that only returned a list of documents, that user was in danger of missing the point entirely. He or she would have read two, three, maybe five reports from the search result of thousands. Since the user did not include "strategic partnership" or "professional services" in the search query the relevance ranking formula probably would not have placed documents rich in those concepts at the top of the search result. Current generation enterprise search engines have the flaw of giving the user what was asked for, not what the user should have asked for had the user already understood the topic.

SUMMARY OF MI ANALYST

MI Analyst combines Northern Light's singular capabilities in content aggregation, integrated search strategy, and powerful indexing technology with text analytics optimized for strategic business research.

MI Analyst is without parallel in the enterprise search industry. It is the only deployed text analytics solution that analyzes large repositories of market intelligence and analyzes the documents on search results not only for statistical relevance to the user's search terms, but for their strategic meaning to the enterprise.

IN CONCLUSION

Northern Light is the leader in strategic research portals for a combination of reasons:

- Due to the scale, sophistication, and research-commitment of our clients, Northern Light is the nexus of best practices for research portals. This gives Northern Light unique insight into what works and what does not work in actual practice.
- Using *SinglePoint*, companies can take advantage of a well-developed suite of portal applications and systems while avoiding the seven-figure investment required to duplicate *SinglePoint's* capabilities. The annual licensing and hosting charge from Northern Light will typically be less than the annual maintenance on a home-grown portal that has the same functionality.
- Northern Light has unequalled content aggregation capability that spans syndicated research publishers, news, scientific journals, financial analyst reports, government databases, Web crawls, and social media.
- Northern Light can integrate internal repositories of research and make it available alongside external strategic research from external parties.
- A company that chooses to host the entire collections of third-party content providers in concentrated directories on the enterprise network incurs significant financial liability for the safeguarding of that material. This liability is avoided if the company alternatively uses Northern Light *SinglePoint* because Northern Light handles all of the content from the third-party suppliers and company employees only download one document at a time when they need to for their research purposes.
- Northern Light *SinglePoint* has a content authorization system that reflects the nature of licensing arrangements for the content. Also, Northern Light is constantly evolving our transactions systems to stay in step as content vendors change the structure of their content licensing plans.
- Northern Light has search technology optimized for strategic business and technology search applications in a variety of important ways, including indexing, syntax options, and relevance ranking.
- *SinglePoint* portals include *MI Analyst*, advanced text analytics optimized for business and technical research that has no peer in the search community. It provides extensive automated tagging of every document in the portal using provided taxonomies of business-relevant concepts. Once tagged and indexed, the concepts can be used to navigate search results and automatically extract meaning.

A key factor is that *SinglePoint* clients have a responsive and flexible partner in the form of Northern Light for research portal best practices, enhancements, and support. We are a focused, boutique shop with a long history of serving the strategic research portal needs of global organizations, and with total control of the *SinglePoint* portal technology and the very powerful and effective search engine that underlies it.

But the best concluding remark is provided by a Northern Light client:

“Northern Light has a great pedigree for search and has a first-class capability — for sure. But, for me, that’s not what it’s about. It’s about getting information to where people can find it, see it, and use it. If the steps it takes to find and see and use are difficult for people, then it’s as if we don’t have the information at all. The information might just as well not exist if it doesn’t get used. [Our company] owns a lot of information — a lot. The question is how do we manage that content to get it to the people who can make the best use of it? That’s the value. That’s our mission. That’s what Northern Light helps us to do.”

So that’s our story. What can we do for you today?

CONTACT INFORMATION

C. David Seuss

CEO

Northern Light

One Constitution Center

Boston Massachusetts, USA, 02129

david@northernlight.com

Mobile: 1-617-515-5771