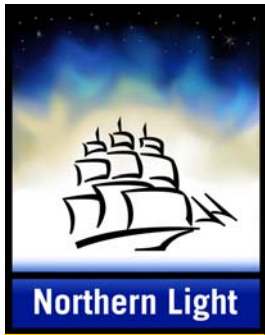


Meaning Extraction for Market Intelligence

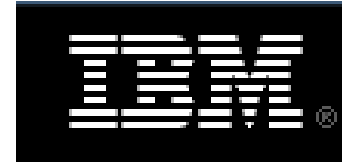
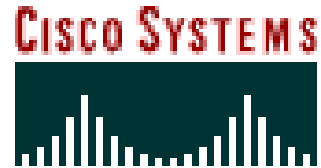
June 2007



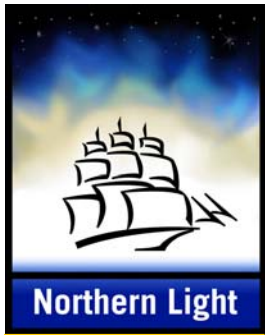


Northern Light

Northern Light hosts strategic research portals for leading companies such as...



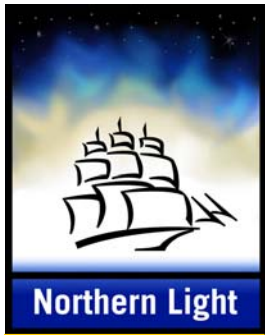
Strategic
Research
Portals



What search engines do

- Ask a search engine to help you find a needle in a haystack, and it gives you a smaller haystack
- Search engines still rely on *the user to process the documents* on the search results,
 - Read them
 - Understand them
 - Relate them to the research problem

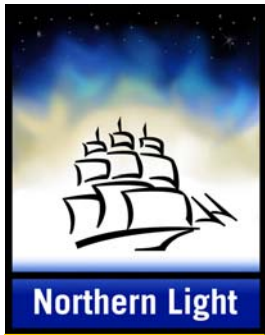




Text analytics to the rescue?

- Useful techniques like
 - Categorization and faceted search
 - Entity extraction
 - Sentiment Analysis
- Text analytics has made a place for itself in reputation management and marketing communications
 - Companies, people, places, products
 - Entity joint occurrence, sentiment scoring, trend analysis
- Not yet penetrating research and MI applications because text analytics in research requires *meaning extraction*

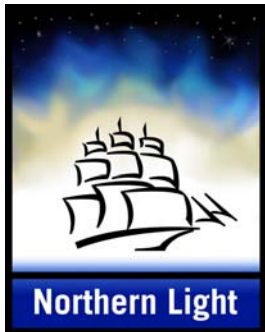




What search engines should do

- What does this results list *mean* to my business?
- In addition to who or what is mentioned, what are the business issues that are identified?
 - What are the trends, the threats, the opportunities?
- If I could read all the reports on the search results, what would I know that I don't know now and might not think to search on?



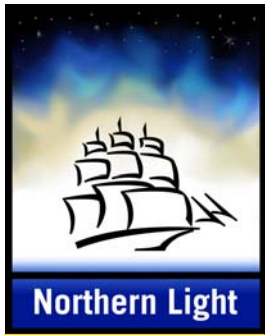


Meaning extraction is the cutting edge

Google Hit Count

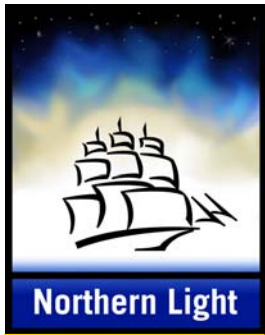
- | | |
|------------------------|---------|
| • “Enterprise search” | 934,000 |
| • “Faceted search” | 256,000 |
| • “Text analytics” | 165,000 |
| • “Entity extraction” | 106,000 |
| • “Sentiment analysis” | 74,600 |
| • “Meaning extraction” | 9,460* |

- * Forty-two of the first 50 Google hits are academic papers, one is a spurious hit on a different topic, one is from [Text Analytic News](#), one is a 2003 article from eContent saying meaning extraction will never work, and five are on Northern Light’s *MI Analyst*



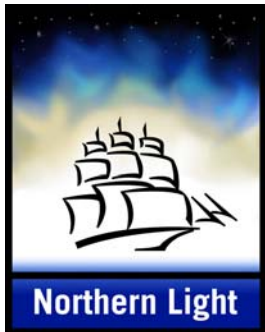
Research and market intelligence documents are different than news stories

- News
 - Short, simple, limited metadata
 - An event, focused
 - Facts, often anonymous
 - Immediate and recent past, short shelf life (weeks, days)
 - Used for reputation management, marketing communications
- Research and market intelligence documents
 - Long, complex, rich metadata
 - Trends, broad
 - Perspective and commentary, authorities
 - Forward looking, long shelf life (months, years)
 - Used to guide the corporate strategy, invent the future



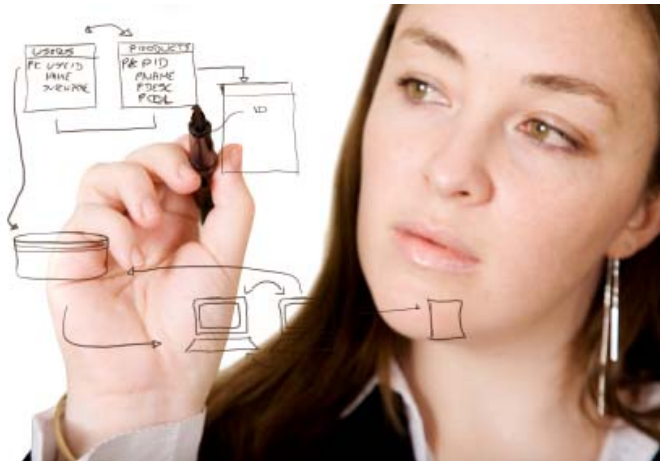
Three stages of text analytics for market intelligence

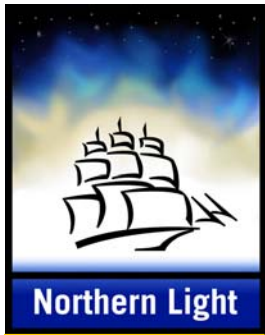
- Traditional: identify the relevant entity classes
 - Companies, people, places
 - Entity joint occurrence, trend analysis, sentiment scoring
- State-of-the-art: identify *meaning-loaded entities*
 - Industry and company customized entity classes: products, markets, technologies, management titles, federal procurement projects, etc.
 - Business issues: market share, falling prices, brand loyalty, market leadership etc.
 - Which reports discuss business issues I care about?
- Next step: identify which combinations of entities in proximity have substantial implications
 - Suggest scenarios of interpretation



Requirements for meaning extraction from market intelligence documents

- Meaning taxonomy
- Document analysis
- Full-text indexing and search
- Proximity

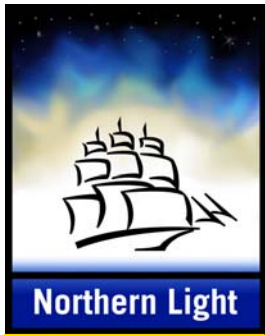




Example market intelligence meaning taxonomy

- + Corporate strategy
- + Cost management
- + Legal and patent
- + Market forecasting
- + Market share
- + Mergers and acquisitions
- + Partnerships and alliances
- + Product marketing
- + Pricing strategy
- + Problem situations
- + Regulatory and compliance

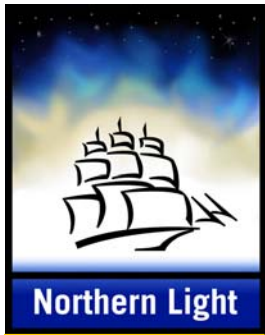




Example market intelligence meaning taxonomy (continued)

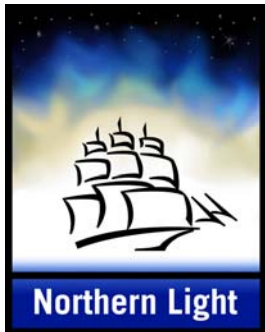
- + Corporate strategy
 - + Competitive strategy
 - + Growth strategy
 - + Strategic positions
 - Market leader
 - Technology leader/innovator
 - Low-cost competitor
 - High-cost competitor
 - First-mover
 - Close-follower
 - Cash cow
 - Rising star
 - Weak business





Document analysis

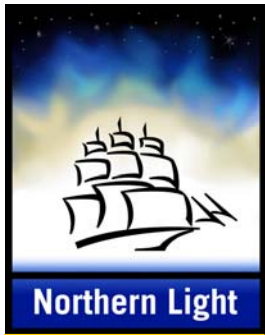
- Basically, a search categorization problem
- Many alternatives: vector, text-string matching, classification queries, and yes, with Web 2.0 leverage, human tagging
- But the essential task is to analyze samples of documents to match them to business issues, and to automate or distribute that process in a production setting
- Concept and thesaurus normalization
 - How many ways can you say “declining prices?”
- Tight integration of the categorization solution with the text analytics and search engine



Search engine

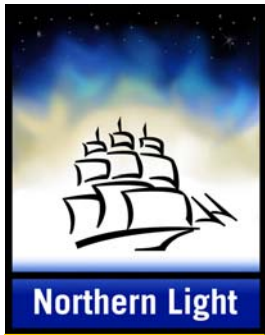
- Full-text index
- Text analytics
 - Facets and entities
 - Business issues
- Proximity
 - Entities to one another
 - Entities to free-text search terms
- Tailor results sort mechanism to the text analytics context
 - By relevance
 - By sentiment
 - By entity density





Scenarios of interpretation

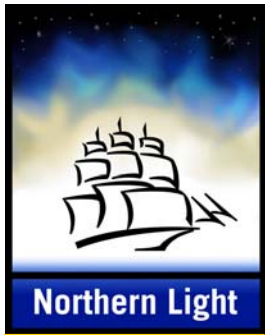
- Rules for suggesting conclusions
- Examples
 - If Company: MegaPath Networks is near Company: Netifice Communications and both are near Business Issue: Competitors, then MegaPath may be a competitor of Netifice
 - If Company: F5 Networks is near Technology: Network Security and is also near Business Issue: Innovator, then F5 may be an innovator in network security
 - If Company: Red Hat is near Business Issue: Declining Prices and near Business Issue: Losing Market Share and near Search Term: Linux, then Red Hat may be losing market share in Linux due to pricing pressure
- The body of rules must be developed by business strategists, not by the search engineers, IT professionals, or librarians on the project



Meaning extraction for market intelligence

- Intersection of disciplines
 - Search, classification, text analytics
 - Document and language analysis
 - Market research and business strategy
- Such teams are hard to form within enterprises and across professional services firms for IT projects





The future of search

- Search engines must evolve to have an *in-depth understanding* of the searched material
- Beyond indexing, classification, facets, and entity extraction, which we all understand by now, the future of search is *meaning extraction*

