

Benefits of Taxonomy: Numbers and Stories and Text Analytics

Tom Reamy
Chief Knowledge Architect
KAPS Group

<http://www.kapsgroup.com>

Author: Deep Text

Agenda

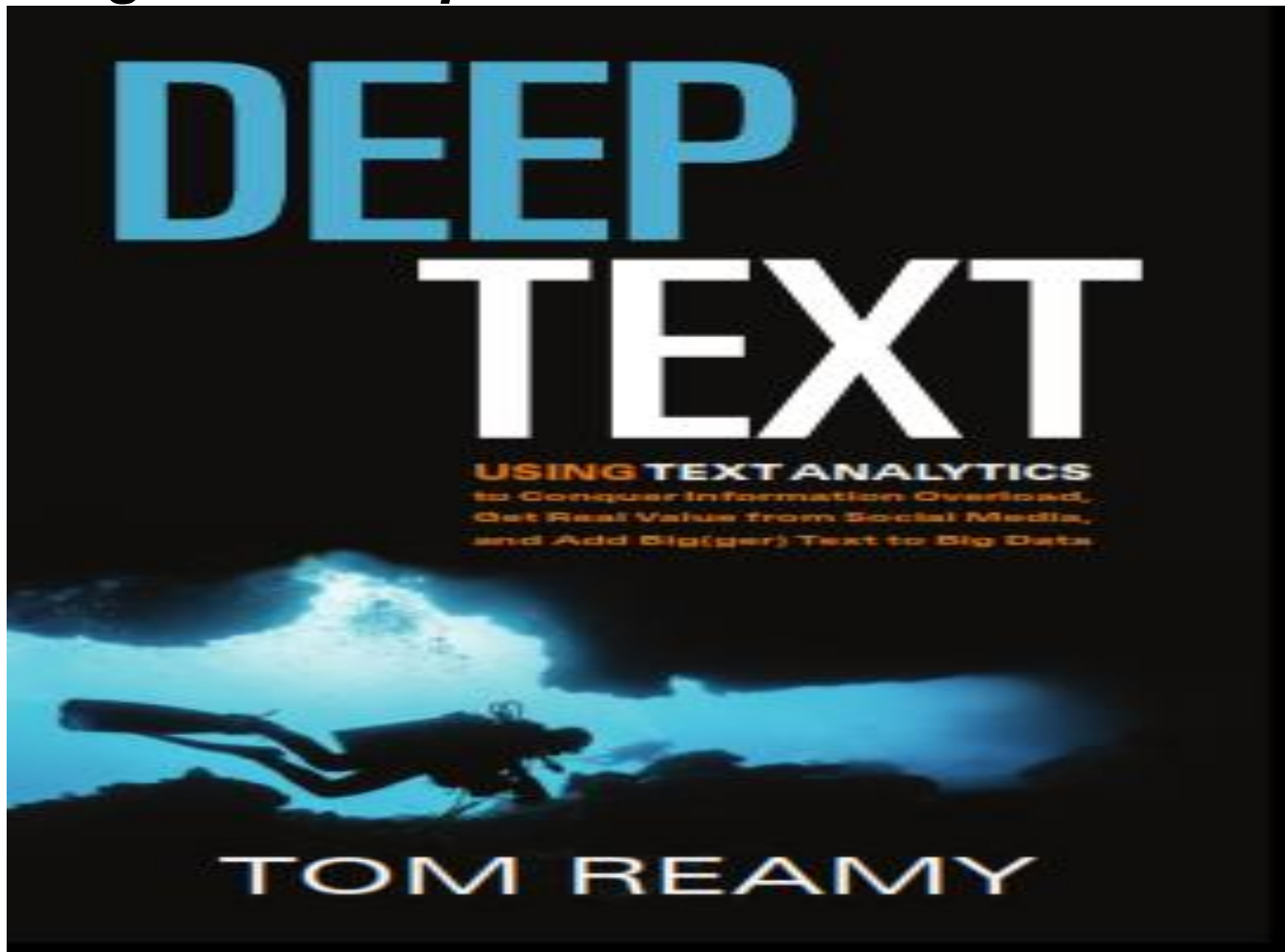
- Introduction
- Benefits of Taxonomies – General Considerations
- ROI Calculations – Good But
- Stories About Taxonomy
- Fatal Flaw?
- Taxonomy in Context
 - Text Analytics
 - Information Platform – CM, Search, etc.
- Selling the Benefits: New Approach
- Questions / Discussions

Introduction: KAPS Group

- Network of Consultants and Partners – “Hiring”
- Text analytics consulting: Strategy, Start-Next level, Development-taxonomy, text analytics foundation & applications
- TA Training (1 day to 1 month), TA Audit
- Partners –Synaptica, SAS, IBM, Smart Logic, Expert Systems, Clarabridge, Lexalytics, BA Insight, BiText
- Clients: Genentech, Novartis, Northwestern Mutual Life, Financial Times, Hyatt, Home Depot, Harvard, British Parliament, Battelle, Amdocs, FDA, GAO, World Bank, IMF, Dept. of Transportation, etc.
- Presentations, Articles, White Papers – www.kapsgroup.com
- Program Chair – [Text Analytics Forum](#) – Nov. 7-8 DC

A treasure trove of technical detail, likely to become a definitive source on text analytics – *Kirkus Reviews*

Book Sign TU Reception – 17:15-18:00



Benefits of Taxonomies

- To Taxonomy or Not To Taxonomy
- That is Rarely the Question
- Unless you are small, a start up, or a high-tech company
- But what about Google, Machine Learning, AI/Deep Learning?
 - Google doesn't work in the enterprise (unless you have a few thousand best bet taggers)
 - Machine Learning clusters are fun for discovery
 - AI works great for perceptions and patterns, not so great for concepts

Benefits of Taxonomies

General Considerations

- Good news is that we don't have to explain what a taxonomy is very often
- Bad news is that there are a lot of very bad taxonomies out there
- Common Situation: Have a Taxonomy
 - Built 5-10 years ago
 - Small and built by amateurs (SME, IT, etc.)
 - Not maintained
 - If used, only for search
 - Top management has never heard of it or used it
- The Question is: How much effort to put into building, applying, and maintaining the taxonomy

Benefits of Taxonomies

General Considerations: No Size Fits All

- Common situation is not that common
- Size variations = different arguments
- Type variations – Publisher, web site, enterprise, government, non-profit, etc.
- Different kinds of taxonomies and applications
 - Subject, actions, motivations, sentiment, etc.
- Not is a taxonomy valuable – It is
 - Is it more valuable than 10 competing information initiatives
- No one argument / approach works for all

Benefits of Taxonomies

Selling the Benefits - Overview

- Start with numerical studies - ROI
- Stories – Pharma example
- Stories – find own real life stories
- Focus of Stories: Business Objectives
- Selling to C Level
 - Different language
 - Need to educate – what it is and why
- Internal Advocacy is Key
- BUT – Fatal Flaw?

Benefits of Taxonomies

ROI Calculations

- IDC study – quantify cost of bad search
- Three areas:
 - Time spent searching
 - Recreation of documents
 - Bad decisions / poor quality work
- Costs
 - 50% search time is bad search = \$2,500 year per person
 - Recreation of documents = \$5,000 year per person
 - Bad quality (harder) = \$15,000 year per person
- Per 1,000 people = \$ 22.5 million a year
 - 30% improvement = \$6.75 million a year
- Other figures double - \$50M

ROI Calculations Limitations

- Problem – no one believes it
- What will staff do with time – extra sip of coffee per search
- Different kinds of search– explore, learn, discover – not find
- Focus on negative – wasted time
 - Positive impact more valuable – greater productivity, better decisions
- Big problem – this is search, not taxonomy
 - Solution could be buy a new search engine
- Assumption that taxonomy will reduce by X% is unproven

Tell a Story: Business Objectives

- Increased richness of K discovery
- Better decisions – complete knowledge
- Analysis – complex issues
 - Purchase complex product - computers, insurance
 - Compare complex decisions – select a college
- Enhanced reporting, more granular and structured reports
- Common nomenclature
 - Bridge across subjects, disciplines, departments
 - Build cortical model – local and global

Tell a Story: Business Objectives New Applications – Social Media

- Understand what customers are saying – satisfaction
 - Customer management effectiveness
 - Insight into customers mind
- Early warning of issues with products
- Lead generation
- Managing brand perception
- Product design insight
- Marketing campaign effectiveness
- Attrition rate management / reduction

Tell a Story: Business Objectives New Applications – Info Apps

- Multiple applications
- Email audit – find money owed
- Products – summary of 700K documents
- Customer support – head off cancelations
- Reduce fraud
- Improve customer support – trends, issues, etc.
- E-Discovery, FOIA
- Range of applications – almost unlimited

Tell a Story: Business Objectives New Applications – Commercial

- Competitive advantage
- Meet regulatory requirements
- New apps – targeted alerts
- Repeat business
- Reputation – can find products, etc.
 - Help find info – save time, money
- Revenue generation - if can't find, leave and don't come back, don't buy

Business Benefits of Taxonomies

Some Flaws in the Story

- Is it the taxonomy or the application?
- Often apps use other types of taxonomies
 - Too often only look at subject taxonomies
- Often apps use other forms of knowledge organizations
 - Ontologies, Knowledge Graphs, Linked Data
- Quality – polyhierarchy, more specific parent of more general term
 - Need professional taxonomists - Difficult to prove
 - Apps return value, apps + good taxonomy more
- Taxonomy (near) Fatal Flaw?

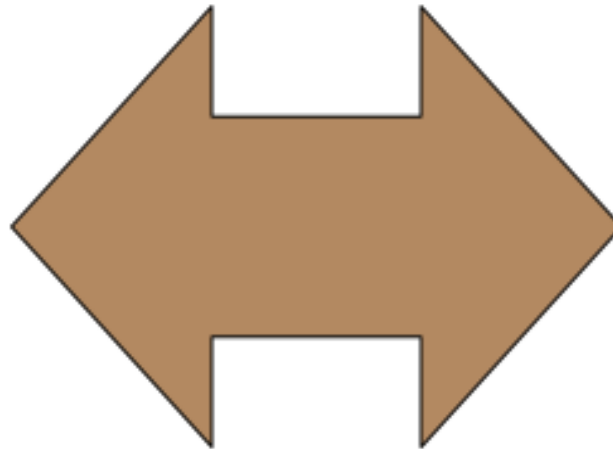
- **Taxonomies are Useless!**

- Taxonomy

- Content

Mind the Gap

■ Taxonomy



■ Content

Benefits of Taxonomies

Fatal Flaw: Mind the Gap

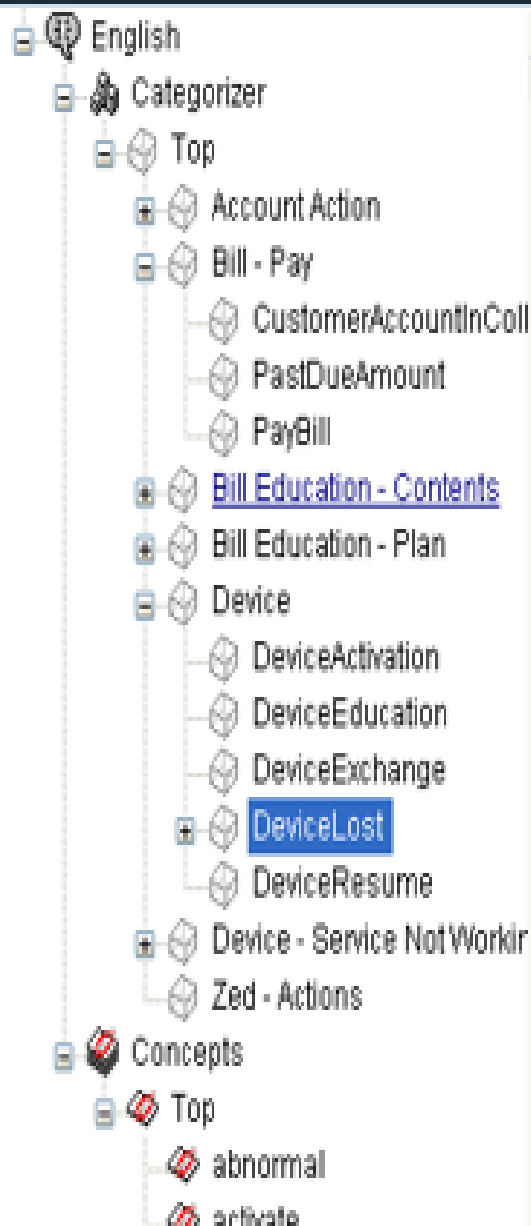
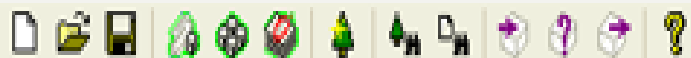
- Tagging documents with taxonomy nodes is tough
 - And expensive – central or distributed
- Library staff – experts in categorization not subject matter
 - Too limited, narrow bottleneck
 - Often don't understand business processes and uses
- Authors – Experts in the subject matter, terrible at categorization
 - Intra and Inter inconsistency, “intertwingleness”
 - Choosing tags from taxonomy – complex task
 - Folksonomy – almost as complex, wildly inconsistent
 - Resistance – not their job, cognitively difficult = non-compliance
- Text Analytics is the answer(s)!

Introduction: Elements of Text Analytics

- Text Mining – NLP, statistical, predictive, machine learning
 - Different skills, mind set, Math & data not language
- Extraction – entities – known and unknown, concepts, events
 - Catalogs with variants, rule based
- Sentiment Analysis
 - Objects and phrases – statistics & rules – Positive and Negative
- Summarization
 - Dynamic – based on a search query term
 - Generic – based on primary topics, position in document

Introduction: Elements of Text Analytics

- Auto-categorization
 - Training sets – Bayesian, Vector space
 - Terms – literal strings, stemming, dictionary of related terms
 - Rules – simple – position in text (Title, body, url)
 - Boolean– Full search syntax – AND, OR, NOT
 - Advanced – DIST(#), ORDDIST#, PARAGRAPH, SENTENCE
- Platform for multiple features – Sentiment, Extraction
 - Disambiguation - Identification of objects, events, context
 - Distinguish Major-Minor mentions
 - Fact Extraction – context around words, concepts
 - Model more subtle sentiment



```
(AND,  
(OR,  
(DIST_5, "[customer]", (AND, "[phone]", "[lost-stolen]")),  
(DIST_5, "[called]", (AND, "[phone]", "[lost-stolen]")),  
(DIST_5, (AND, "[customer]", "[called]", "[lost-stolen]"))  
),  
(NOT,  
(OR, "[activate]", "[swap]",  
(DIST_5, (OR, (OR, "[customer]", "[called]"), "[lost-stolen]"), "[restrict]"))  
)  
)
```

Benefits of Taxonomies: Content Models Adding Structure to Unstructured Content

- Documents are not unstructured – variety of structures
 - Sections – Specific - “Abstract” to Function “Evidence”
 - Corpus – document types/purpose
 - Textual complexity, level of generality
- Content Model – taxonomy of document types / Sections
- Text Analytics – categorization rules weight sections
- Good taxonomy + text analytics + content models = 98%
- Clusters and machine learning – at section level, not document
- Future = Combine machine learning and rules

Boehringer Pilot One Drug Names Disease

English

Categorizer

Top

Diseases

arthritis

Benign Prostatic Hyperplasia

Cancer

Hypertension

Deep Vein Thrombosis

HIV

Pulmonary Disease

Drug Names

afatinib

clonidine

dabigatran

meloxicam

tamsulosin

telmisartan

tiotropium

Concepts

Top

BI Drugs

Diseases

arthritis

```
(OR,  
  _/article/title:"[arthritis]",  
  
  (AND, _/article/mesh:"[arthritis]",_/_article/abstract:"[arthritis]"),  
  
  (MINOC_2, _/article/abstract:"[arthritis]"),  
  
  (START_500, (MINOC_2,"[arthritis]"))  
)
```


Benefits of Taxonomies

Mind the Gap

- Automatic tagging?
 - Fatal Flaw = accuracy, brittle – new content, effort
- Hybrid Model: Combination of machine and human
 - Publish Document into content management
 - Text Analytics analysis -> suggestions for categorization, entities, metadata
 - Present to author / editor
- Cognitive task is simple -> react to a suggestion instead of select from head or a complex taxonomy
- Feedback – if author overrides -> suggestion for new category
- Scale – All content or selected content

Benefits: Selling the Vision

- All of that is a complex sell – how to do it?
- New Approach – Mini-Demo
- One approach – Any size
 - One week to 2 Months (Demo to POC)
- Elements
 - Taxonomy (Old, one branch) – 10-20 nodes to 100
 - Sample content – 10-20 documents per node
 - Simple content model – document sections
- Build categorization rules for all nodes
- Demo – Simple search (15%-50%) to 90%+

Benefits: Selling the Vision

Mini-demo

- Something that people can see, touch, play with
- Real application with real content
- See the value of Taxonomy + Text Analytics
- Appeal to all audiences – Librarians to KM to technology geeks to executives
- Option – Comparison with fully automatic clusters
- Start of building a foundation for full enterprise
 - Full POC can build (most of) that foundation

Benefits of Taxonomies

Conclusions

- No one size fits all – adapt message to organization
- Numbers, Stories, Mini-Demo
- ROI – Wrong Question – What is the ROI for organizing a company
- Fatal Flaw of taxonomy – Mind the Gap
- Text Analytics powering a hybrid tagging minds the gap
- Catonomy is (part of) the answer
- Taxonomy (and text analytics) are major components of the semantic infrastructure
 - Taxonomies new and old types, ontologies, content management, search, auto-categorization and entity extraction, sentiment, Info Apps
- Sell the benefits – all of the above

Questions?

Tom Reamy
tomr@kapsgroup.com

KAPS Group

Knowledge Architecture Professional Services

<http://www.kapsgroup.com>

AI and Taxonomy

- Relational Frame Theory - RFT
 - Coordination – (similarity) dog is same as hound
 - Distinction – (difference) – white dog different than a black dog
 - Opposition – a black dog versus a white cat
 - Comparison – this dog is bigger than that dog
 - Spatial – this dog is on the left
 - Temporal – I fed the dog before the cat
 - Hierarchical – a dog is a sort of mammal
 - Causal – a dog bite causes me to cry