

# Artificial Intelligence and Taxonomy

Tom Reamy  
Chief Knowledge Architect  
KAPS Group

<http://www.kapsgroup.com>

Author: Deep Text

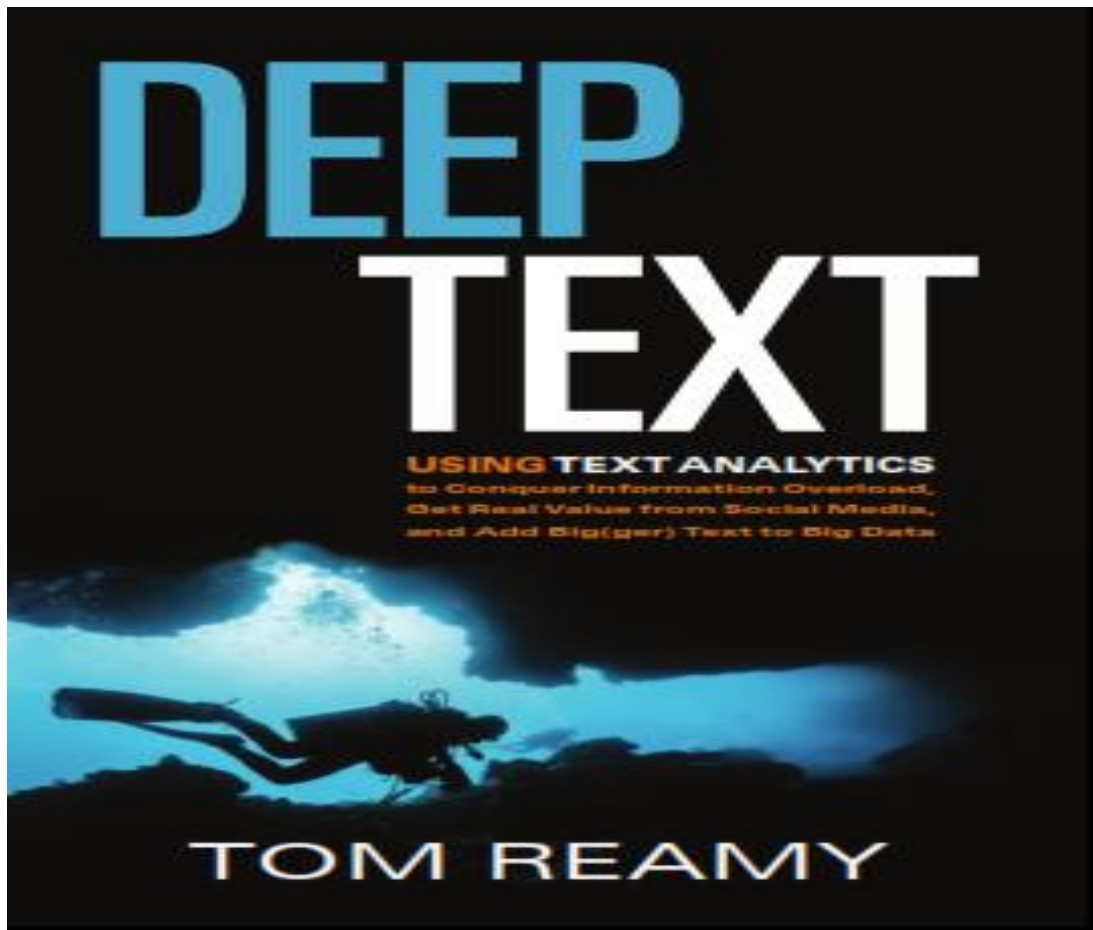
## Agenda

- Introduction
- AI: Past and Present
- AI and Taxonomy – 2 Way Street
- Taxonomy and Knowledge Organizations
- Future Directions
- Conclusion

## Introduction: KAPS Group

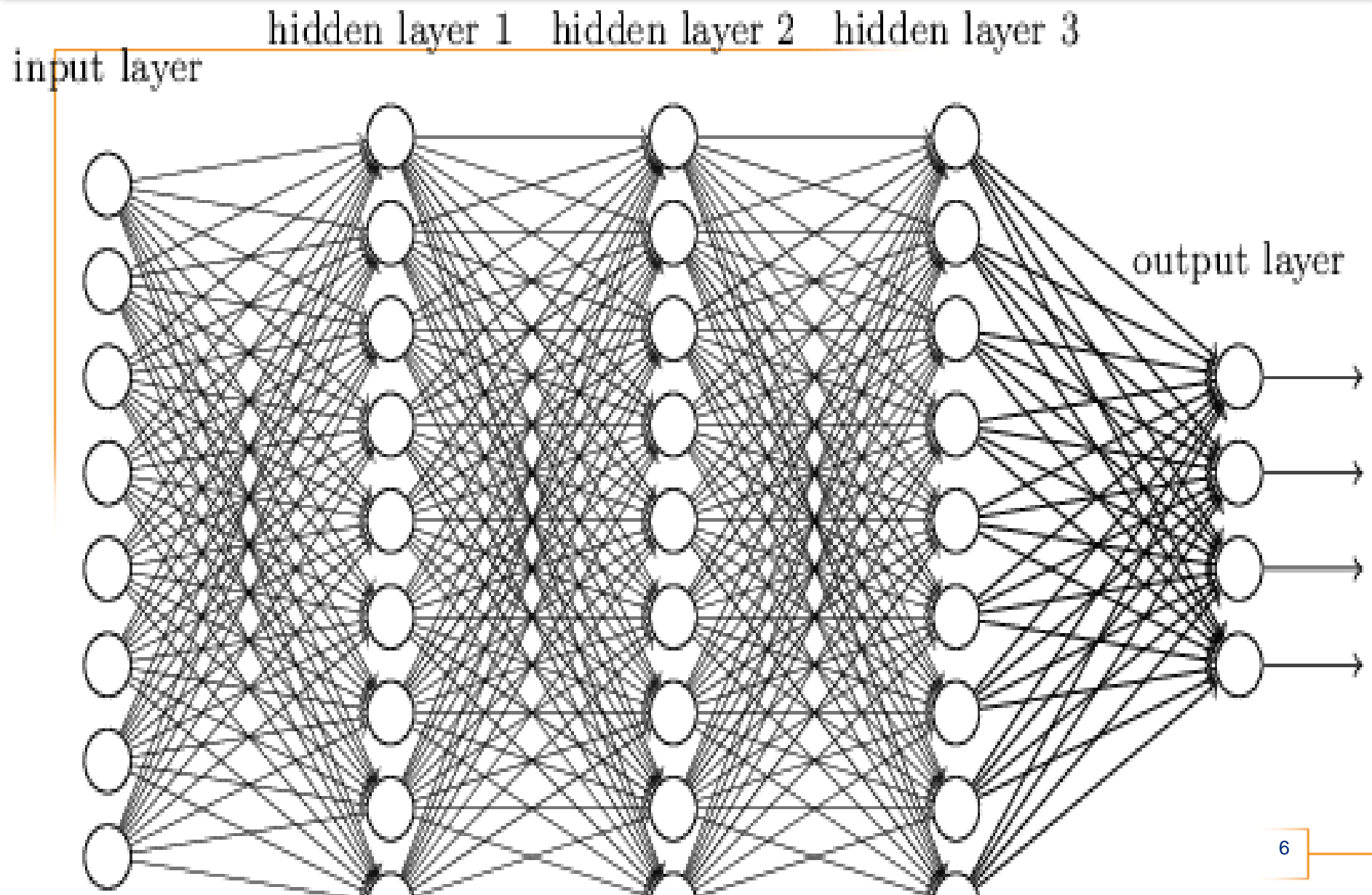
- Network of Consultants and Partners – “Hiring”
- Text analytics consulting: Strategy, Start-Next level, Development-taxonomy, text analytics foundation & applications
- TA Training (1 day to 1 month), TA Audit
- Partners –Synaptica, SAS, IBM, Smart Logic, Expert Systems, Clarabridge, Lexalytics, BA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

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



## AI Predictions

- By 2045, AI will take over 80-90% of librarianian jobs!
- Let me tell you a story.
- Early start 1980's to first AI winter – 2 approaches
  - Symbolic AI – rules / human models – too rigid, costly
  - Neural Networks – ML – too stupid
- New hype is Deep Learning – neural nets on steroids
- New = size and speed = larger and multiple networks
- Strongest in areas like image recognition
- Weakest – concepts, subjects, deep language, metaphors, etc



## AI and Taxonomy

### Deep Text vs. Deep Learning

- Deep Learning is a Dead End - accuracy – 60-70%
- Black Box – don't know how to improve except indirect manipulation of input
  - Watson – “We don't know how or why it works”
  - Susceptible to bias – hard to fix
- Domain Specific, tricks not deep understanding
- Needs millions of examples – humans jump to conclusions
- No common sense (things fall, don't wink in and out of existence)
  - No strategy to get there (faster not enough)
- Don't believe (all) the hype

## **AI and Taxonomy**

### **Two Way Street**

- AI for Taxonomy
  - AI for Building Taxonomies
  - AI Adding New Knowledge Organizations – taxonomy+
- Taxonomy for AI
  - Taxonomy of AI concepts
  - Taxonomy can add structure (conceptual and linguistic) to AI
- Current best use – Text Analytics

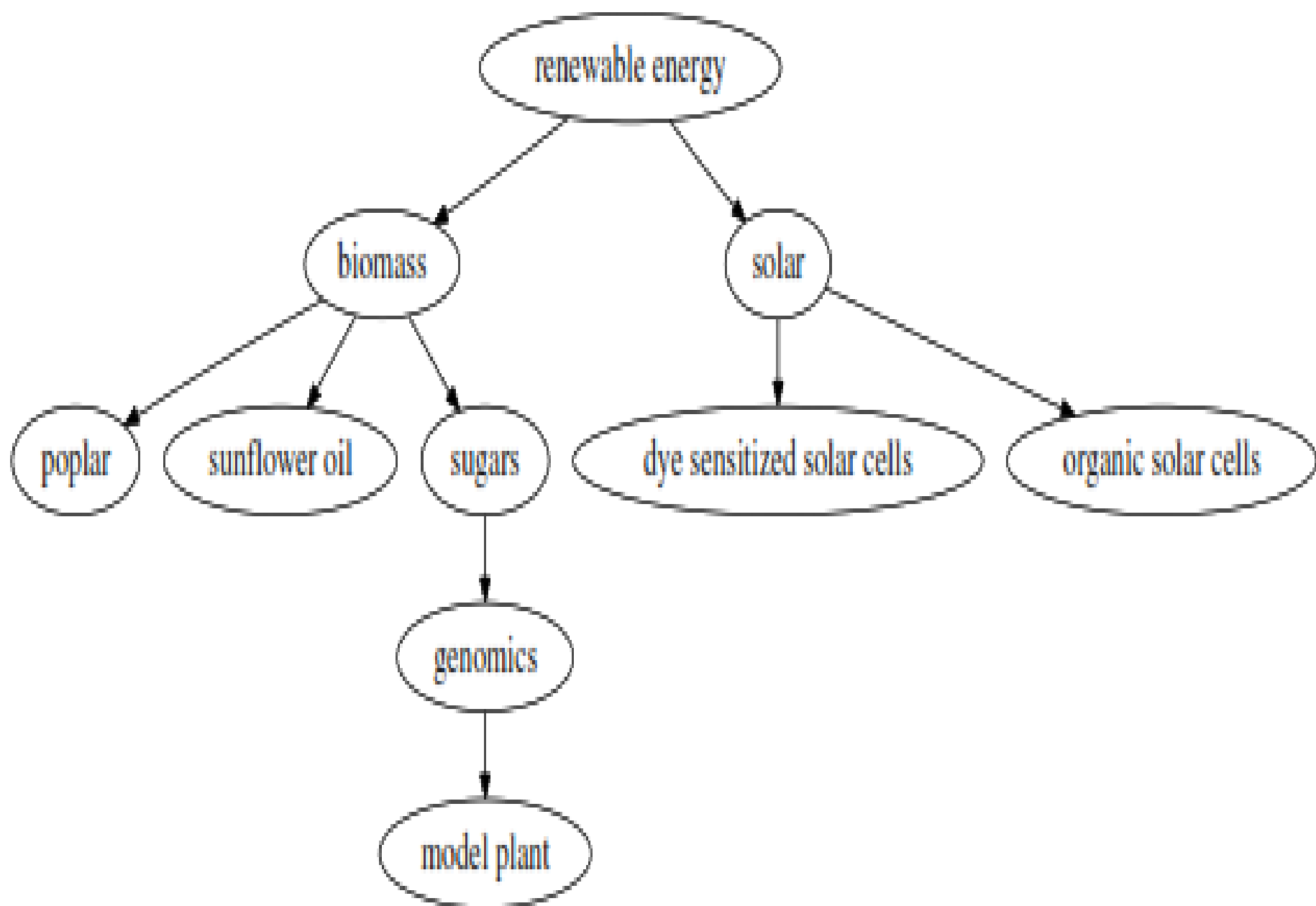


## AI and Taxonomy

### Automatic Taxonomy

- Most Text Analytics vendors offered – very poor results, dropped
- New techniques – getting better but don't give up your taxonomist day job
- “Automatic” – but not a taxonomy – cluster of co-occurring terms
  - Suggest terms and relationships
- Hybrid – text mining on steroids
- “Automatic” – huge human effort to design approach, mathematics, select content, seed taxonomies, keyword selection, data prep – then voila!



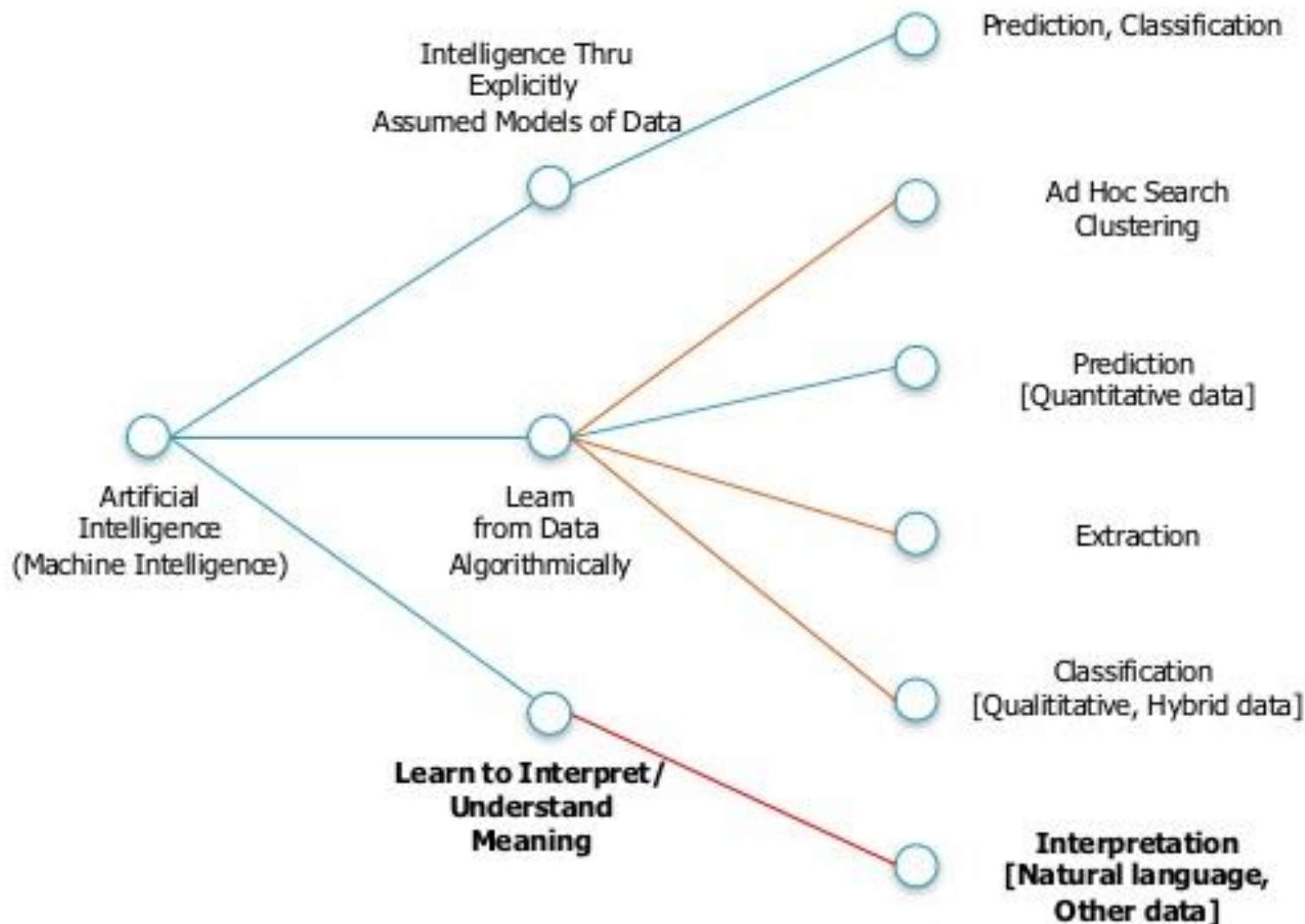


# Taxonomy and AI

## Taxonomy of AI

# AI in the Enterprise

## A Taxonomy of Machine Intelligence Problem Types



## Taxonomy and AI

- Taxonomy can add structure (conceptual and linguistic) to AI
- Multiple types of Knowledge Organization
  - Taxonomy – concepts, hierarchical
  - Ontology – any type of relationship, things and concepts
  - Knowledge Graphs – triples, unlimited, no overall structure
- K Graphs and hierarchical – best way to merge?
  - Modules, facets
  - Hierarchical network models
- Catonomy – taxonomy + categorization rules
- New types – cognitive science - RTF, other?

## 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 bit causes me to cry

■

## AI and Taxonomy Conclusions

- AI is still in the land of hype
  - For language, concepts, taxonomy
  - Data dimension means no new AI winter
- The future is not Deep Learning – need advanced KO
  - Neural nets only associative, need more like RFT
- AI can be useful in building taxonomies – suggestive
- Future is Integration of AI/Deep Learning and Deep Text rules
  - Integration platform is text analytics
- For more – Text Analytics Forum – Nov. 8-9 DC and Taxonomy Boot Camp – Nov. 6-7



# Questions?

:

## Resources

### ■ Books

- Deep Text: Using Text Analytics to Conquer Information Overload, Get Real Value from Social Media, and Add Big(ger) Text to Big Data
  - Tom Reamy
- Women, Fire, and Dangerous Things
- Don't Think of an Elephant
  - George Lakoff
- Knowledge, Concepts, and Categories
  - Koen Lamberts and David Shanks
- Thinking Fast and Slow
  - Daniel Kahneman
- Any cognitive science book written after 2010

## Resources

- Conferences – Web Sites
  - Text Analytics Forum - All aspects of text analytics
    - <http://www.textanalyticsforum.com>
  - Semtech
    - <http://www.semanticweb.com>
  - Dataversity Conferences
    - <http://www.dataversity.net/>
  - Sentiment Analysis Symposium
    - [www.sentimentsymposium.com](http://www.sentimentsymposium.com)

## Resources

- LinkedIn Groups:
  - Text Analytics
  - Text Analytics Forum
  - Taxonomy Community of Practice
  - Sentiment Analysis
  - Text and Social Analytics
  - Metadata Management
  - Semantic Technologies, Semantic Web
  - Association for Information Science & Technology

## AI and Taxonomy

### Taxonomy to Text Analytics Rules

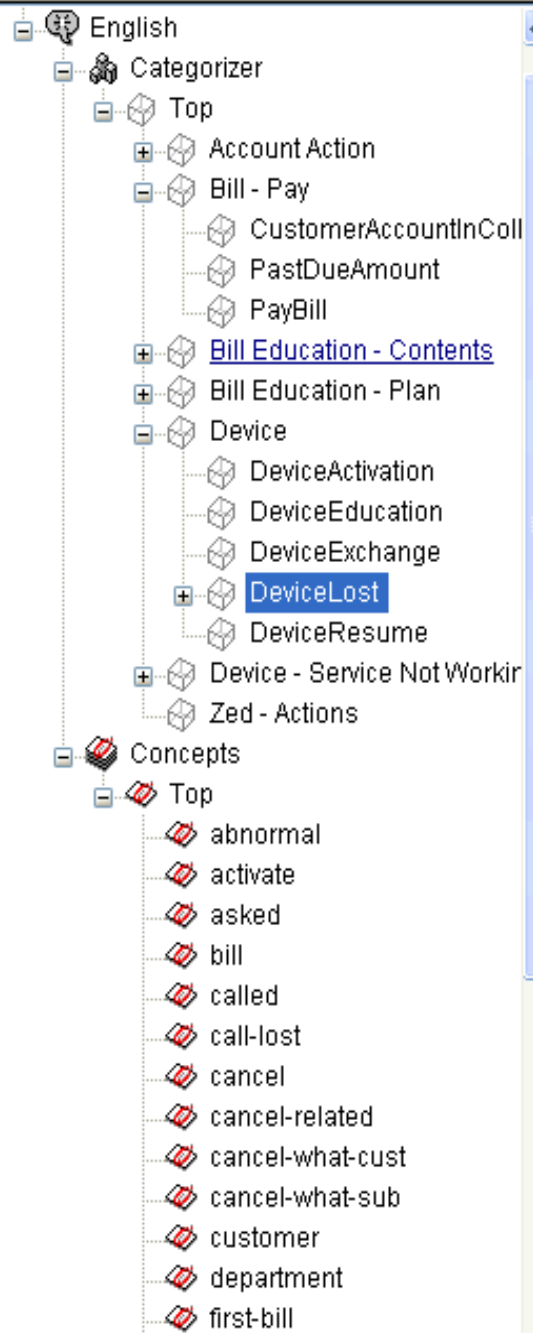
- Create a set of rules from the taxonomy
  - Concepts, synonyms, semantic distance, etc. – Synaptica
- Create a set of rules based on training documents
  - SME select good sample documents for each taxonomy node
  - Techniques
    - Pull out important and unique terms
    - Create a statistical signature that can be used to classify new documents
  - Limited – very content specific – can't build on
- New techniques - Multiple Modules – combine or competition
  - Auto-rules from text – Poly-structure – clusters to sections

## **Introduction: Elements of Text Analytics**

- Text Mining – NLP, statistical, predictive, machine learning
  - Different skills, mind set, Math & data not language
- Semantic Technology – ontology, fact extraction
- 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

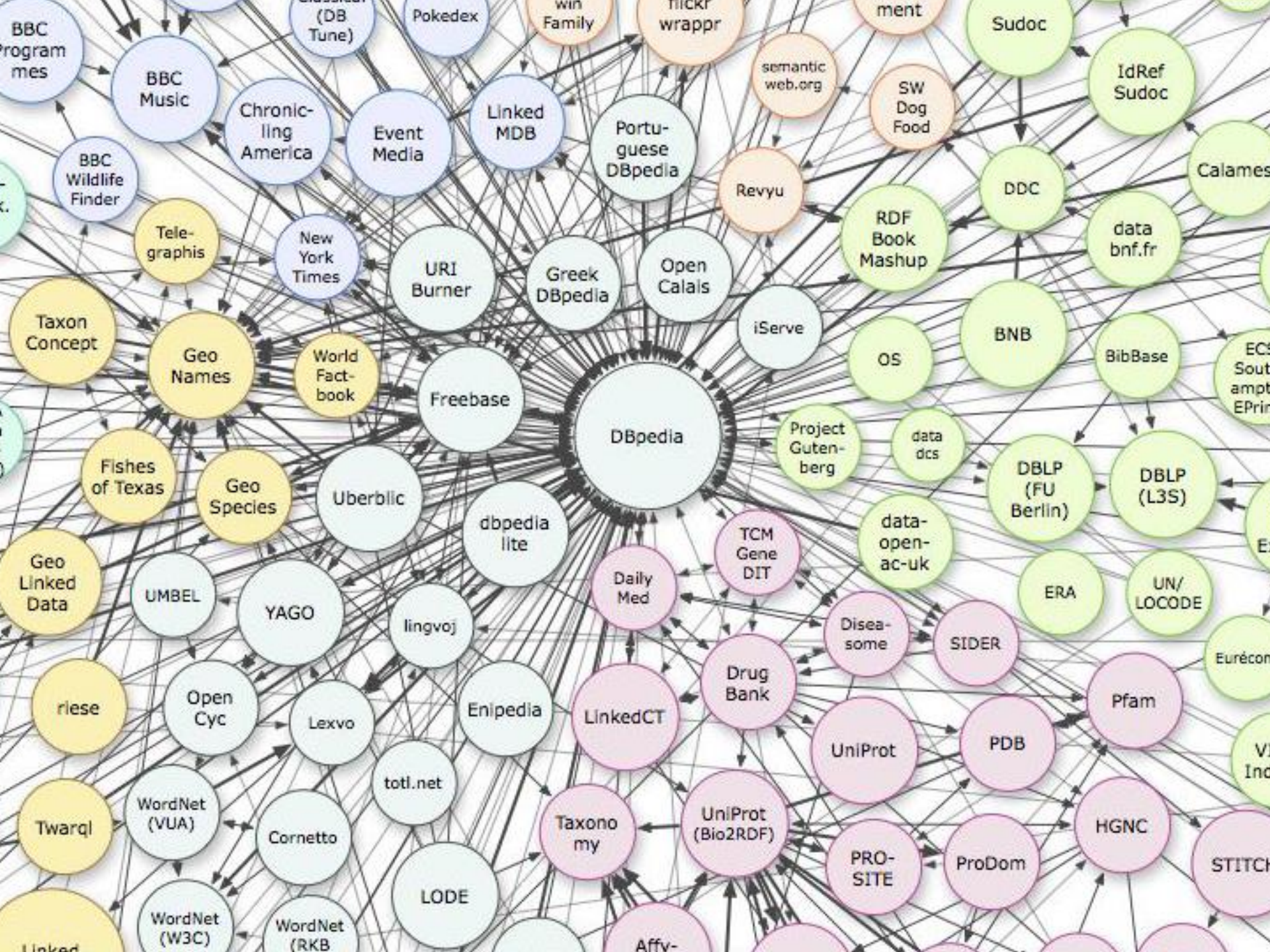
## Introduction: Elements of Text Analytics

- Auto-categorization
  - Training sets – Bayesian, Vector space
  - Terms – literal strings, stemming, dictionary of related terms
  - Rules – simple – position in text (Title, body, url)
  - Semantic Network – Predefined relationships, sets of rules
  - 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
  - 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]"))
)
)
```





## Deep Text Reviews

- “A treasure trove of technical detail, likely to become a definitive source on text analytics” – *Kirkus Reviews*
- “A remarkably authoritative deep-dive into a field that will be brand new to many and eye-opening for all.” – *Kirkus Reviews*
- “One of the main strengths of this book, though, is that even when its content is highly technical, it’s so well-organized and tightly written that it’s quite enjoyable to read.” – *Kirkus Reviews*