

Text Analytics Forum 2025

Program Chair: Tom Reamy
Chief Knowledge Architect
KAPS Group

<http://www.kapsgroup.com>

Author: Deep Text
tomr@kapsgroup.com

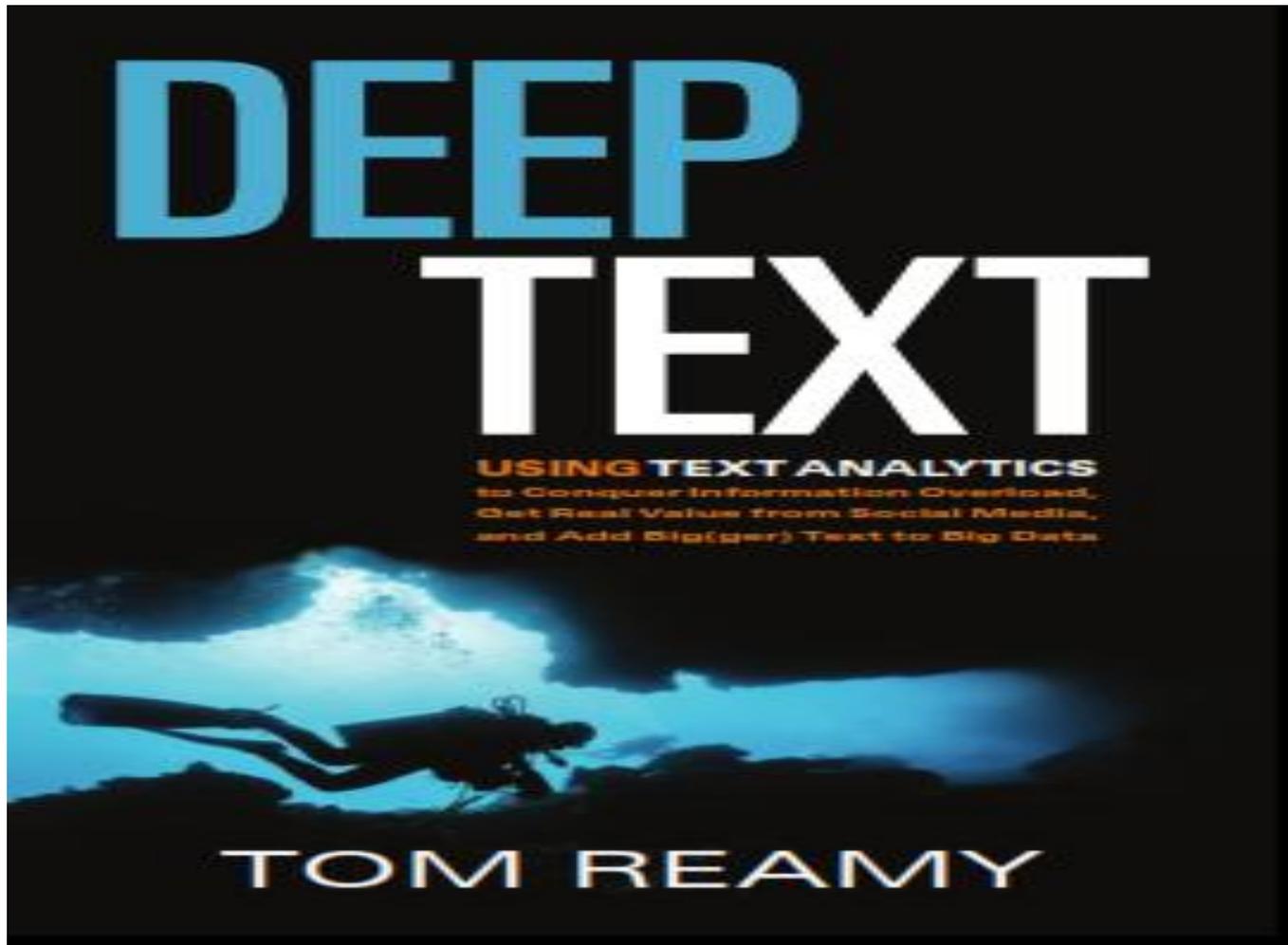
Agenda

- Introduction
- Conference Highlights and Themes
- Overview of Text Analytics
 - Categorization, Data Extraction, Supplemental
- Gen AI
 - AGI in Two Years OR The Bubble Bursts OR?
- Questions

Introduction: KAPS Group

- Network of Consultants and Partners - 2002
- Text analytics consulting: Strategy, Development-taxonomy, text analytics foundation & applications
- Mini-Projects – get started or take to next level
 - Strategy, Mini-POC - Categorization
- Partners – Semantic Arts, Expert AI, Megaputer, SAS, Smart Logic-Progress, Lexalytics, BA Insight, BiText, Synaptica
- Clients: Genentech, Novartis, Northwestern Mutual Life, Financial Times, Hyatt, Home Depot, Harvard, British Parliament, Battelle, Amdocs, FDA, GAO, World Bank, IMF, IFC, Dept. of Transportation, RWJF, Intel, Kellogg Foundation, Foundry (IDG), Service Now, etc.
- Presentations, Articles, White Papers – www.kapsgroup.com
- Program Chair – [Text Analytics Forum](#)

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



Text Analytics Forum (TAF) Introduction Conference Highlights and Themes

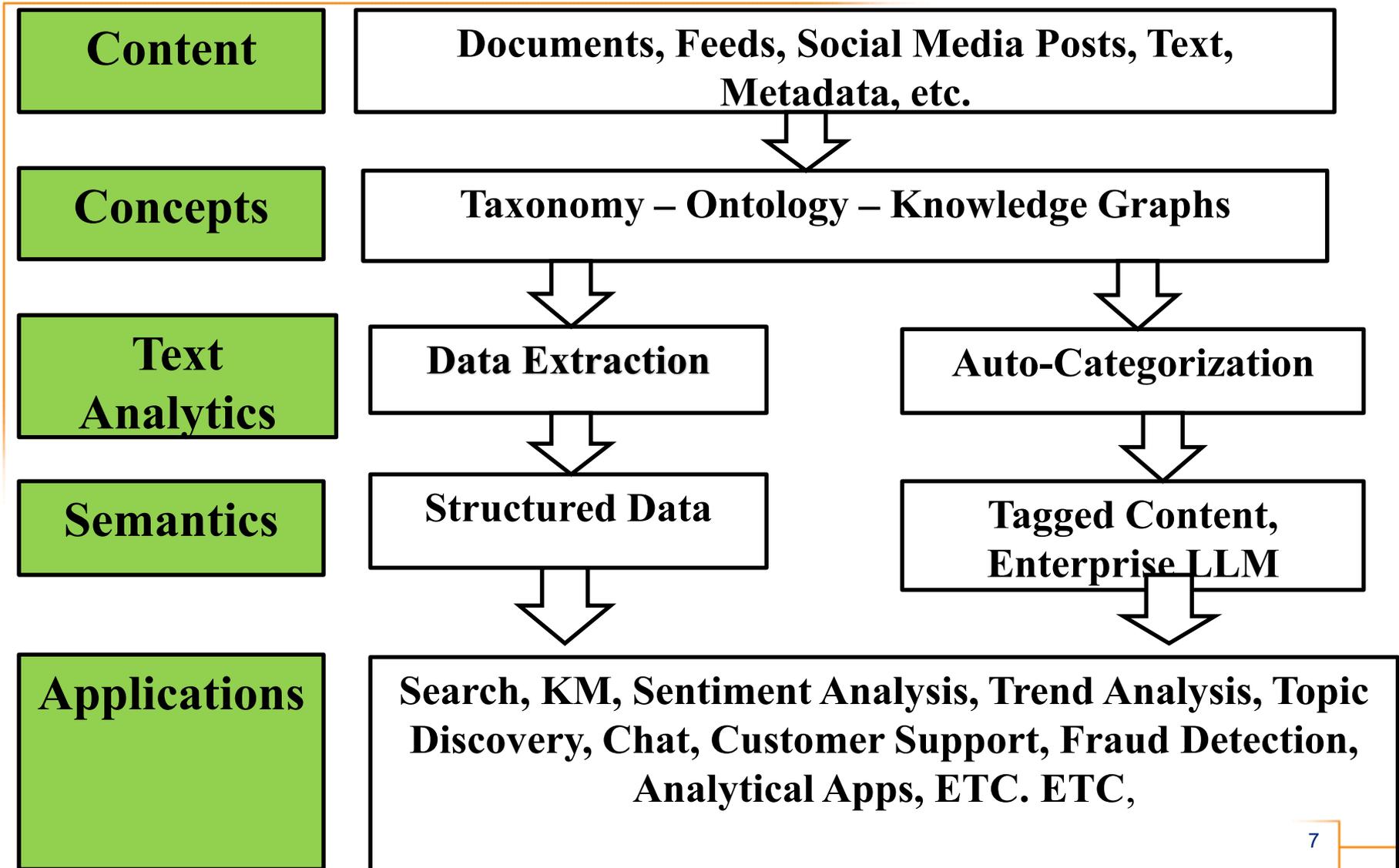
- Theme – Good Foundations Make Good Applications
 - Gen AI: AGI or Bubble Burst
 - RAG – better, but GIGO
 - TA to the rescue – accuracy, curated content
- Where-Meeting Room Level – Russel-Hart-Cannon/Rayburn
- Wednesday –Hybrid Solutions – Text Analytics and Gen AI
 - Development and Applications
 - Adding Structure – Taxonomy, Knowledge Graphs, Content Models
- Thursday
 - Building a Text Analytics – Gen AI Application Platform
 - Use Cases and Tools and Techniques
 - Customer Insights & Social Media
 - Knowledge Graphs

Text Analytics Forum (TAF) Introduction

Elements of Text Analytics

- Auto-categorization - Statistical – Rules
 - Brains of the outfit: (Rules) Makes everything else smarter
 - Sentiment Analysis – positive & negative, attitudes
 - Advanced - racial equality, social media analysis
- Data Extraction – entities, concepts, events, facts
 - Analytical applications, enhance search - facets
- Supplemental – NLP, summarization, variety of analytical
 - NLG – Language generation
 - Text Mining – NLP, feed analytical apps, terms for categorization
 - CSM – Content Structure Model – not Bag of Words

Text Analytics and Gen AI in the Enterprise



Text Analytics Forum (TAF) Introduction

Gen AI Hype

- History of AI – story of hype and failure, boom and bust
- AI Hype – both sides guilty – AI will kill us all, AI will transform everything
- “Soon AI models could perform the entire scientific method, without humans” – Time
- “Meanwhile, driverless vehicles will put truck, bus, and taxi drivers out of work.” -Time
- “We are at the beginning of the agentic era, the most significant transformation in the history of work.” – Marc Benioff
- “We’re creating systems that understand text, voice, images, and code,…” – Marc Benioff

Text Analytics Forum (TAF) Introduction

Generative AI and Text Analytics

- Gen AI doesn't understand anything.
 - Text prediction on steroids`
 - We project intelligence on to almost anything – Eliza on
- General limitations
 - Hallucinations – esp. critical areas – finance, legal, etc.
 - Security-bias, lack of transparency
 - Limits of scale – diminishing returns – billions = 3-5% increase
 - Cost – need for 100's billions documents, banks of super computers
 - 85% of Gen AI projects fail to deliver significant value?
 - 77% workers – AI increased their work load, not productivity
 - Need humans in the loop – but this limits and slows AI

Text Analytics Forum (TAF) Introduction

Generative AI and Text Analytics

- Issue of quality
 - New advanced version – can't count to three
 - Can't learn on its own
 - Correlation – no understanding of causality
 - LLM Brain rot: Esp. if trained on clickbait, short video social media
 - 75% of new web content is at least partially generated by Gen AI
- Negative Impacts
 - Higher AI literacy brings more overconfidence
 - Greater use of Gen AI = lower critical thinking skills
 - “The danger is not that machines develop too much intelligence, but that we stop exercising our own.”

Text Analytics Forum (TAF) Introduction

Gen AI AGI or Bubble Burst

- AGI? Not a chance without new approach, breakthrough
 - Yann LeCun-world models
- Which AI – Gen, ML, Behavior Prediction? None/all of the above
- “We must choose wisely” – Marc Benioff
- History tells us that some will choose wisely, some will choose badly.
 - Most hype fails to deal with those people, organizations, and countries that will use AI for bad purposes”
- Agents based on current LLMs – danger of low quality and catastrophic errors – Chinese automated attack with Claude based agent

Text Analytics Forum (TAF) Introduction

Generative AI and Text Analytics

- How add quality – TA, ML, Gen AI together
- How do you combine TA and Gen AI?
- General – Gen AI provides general answer, TA adds precision
- Example – Search - Hybrid and multi-path solution
 - Recall – ML and Gen AI
 - Precision – Auto-cat
- RAG – 3 parts
 - Answers, List of Links, Gather answers, ask for executive summary
 - Answers & Links – only as good as your search – TA makes better
- What do you get by adding TA? Accuracy! Which impacts all attempts to utilize and analyze documents

Text Analytics Forum (TAF) Introduction Gen AI and Semantic AI (TA)

- Add curated content to existing LLM or prompt
- Issue – quality of content – cost of getting good content
 - Statistical – there is more bad content than good
 - Human curation is expensive
 - Search engines not accurate enough
- Answer – semi-automatic content curation – auto-categorization
- Better training data – extract more out of each document – Use CSM
- Prompt engineering - queries and embed content in prompt
- Using agents that aren't based on a good foundation– disaster?
- Future: Deeper integration of Semantic AI, ML, and Gen AI – and ?
 - Combine world models (perception) and language models

Text Analytics Forum (TAF) Introduction

Conclusions

- Gen AI Boom or Bust? Bust: money, but not useful tool
- Main obstacles are complexity of language, scale, and no physical component - causality
- Whole lot of projecting going on
- Danger - Putting up with answers because “automatic” or because its Microsoft, Google, Amazon, or OpenAI
 - Gen AI is making us stupid
- TA makes Gen AI Smarter and Safer
- No such thing as unstructured text – Content Structure Models
- Future is Integration of GPT/LLM and Deep Text rules + world models
 - Integration platform is text analytics
- Not AI or not AI, but how it is developed and used – world models