

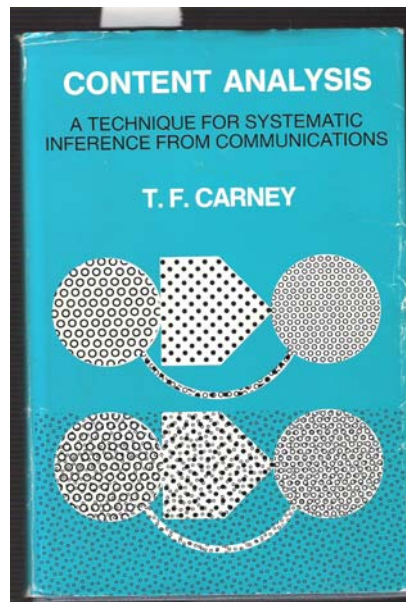
Using faceted classification to capture intelligence from news flows

Presentation to ISKO UK, University College, London, November 5, 2007

Jan Wyllie and Simon Eaton, Open Intelligence Ltd. Email - j.wyllie@open-intelligence.co.uk

Roots and Covers

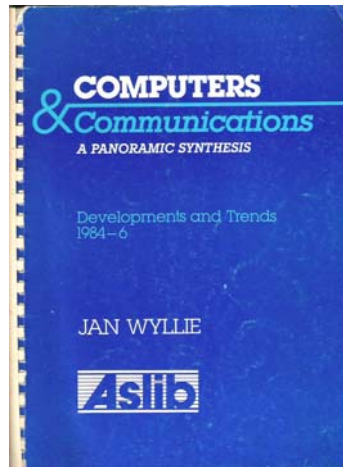
- I came to the subject of faceted classification from a very different route from most people here. So I think it is important for people to know something about how I come to be here.
- After a career in journalism and as a globe trotting officer in the Canadian Department of External Affairs, in the 1970s, I learned the content analysis research methodology at the **Canadian Trend Report** in 1979. We used the methodology to compile reports and provide two-day workshops to top level decision makers in government and industry. I recall they paid a great deal of money for the privilege. I became editor of a new category (which I had suggested) called Communications in which we included all the sexy new high tech computer and comms stuff. As part of our employment contract we had to sign a document saying we could not ever use what they called a proprietary research methodology elsewhere on pain of being sued for every penny we had.
- For the first two years after I returned to the UK for romantic reasons, I commuted every six months from London to Montreal to edit Communications. It was during this period that a student friend of mine in London found **Tom Carney's** 1972 book, here at UCL in the Senate Library. What joy and release! Content analysis could *not* be proprietary. It had a long history. And I had the best book about the subject which hardly anybody knew about. Here are just a couple of introductory quotes.



•
 Quotes from book

- *“Content analysis is a collection of techniques which improve the quality of inferences made in the study of communications – whether they are written verbal or even pictorial. A well thought through system of analysis makes possible the posing of different questions with some assurance that the answer will fairly represent the material.”*
- *“If we are to focus unwaveringly, we need discipline. This is what content analysis imposes. It forces us to be very conscious about just what we are looking for, and why we are looking for it – about what is sometimes called our frame of reference. It also forces us to hold this frame of reference steadfastly. Content analysis is a way of asking a fixed set of questions unfalteringly of all of a predetermined body of writings, in such a way as to produce countable results.”*
- Since the 1920’s content analysts have used what they call multivariate classification schemas to study large collections of text, images or objects. Since then it has been widely used in Western intelligence agencies and the social sciences. One famous war time project could identify German troop movements by analysing German train timetables in the context of local press reports announcing changes. These days, it has been taken over by computers counting instances of words in natural text with the purpose of identifying styles or social assumptions. This is called quantitative content analysis.
- We practice qualitative or ‘theoretical’ content analysis using multivariate classification schemas. As far as I can see, both multivariate analysis and facet analysis mean much the same thing. One possible difference is their purpose. Ranganathan seemed to want a comprehensive, unchanging way of identifying fixed information objects with his famous five fundamental categories. Content analysis varies its facets depending on the purpose of the research. It often concentrates on change – events, actions, trends.
- So in 1986, I founded Trend Monitor with the intention of applying content analysis to the computing and communications fields, both as subject matter and as a useful tool. My first independent work, **Computers and**

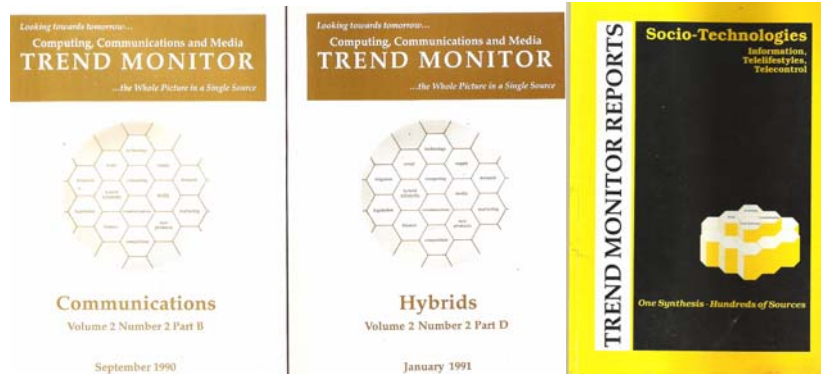
Communications, A Panoramic Synthesis, was paid for by Monty Hyams founder of Derwent, edited by his son Peter (who went on to edit Information World review, and was sold out ... by Aslib. It used content analysis and paper-based hypertext



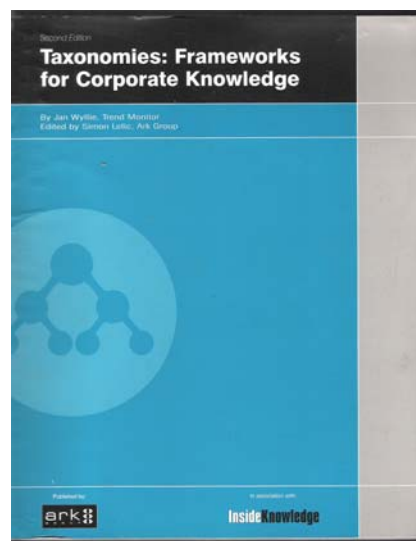
- About that time, I also met **Tony Kent**. It was the start of a seven year collaboration brought to an end by his untimely death. He taught me about ‘information science’ – a term he did not like – and how the software of text databases worked. We used his text database software, STRIX on applying meta-languages and clever counting systems. We had lots late night discussion and some whisky too.



- After the success of the **Panoramic Synthesis**, we did joint venture deal with **Aslib** to produce **Trend Monitor Reports**. The deal negotiated in the pub with Richard Coleman, Aslib’s publisher was: We paid for the research and writing. Aslib paid for the marketing and fulfilment. The revenue was split 50-50. It was the dumbest deal I ever did! Nearly all Aslib’s money went on printing and distribution, virtually none on marketing. Nevertheless, we did gain some great and loyal customers. Liz Orna was our first and perhaps our last. Note how *Hybrids* evolves into *Socio-Technologies*



- We were then taken over by **Spikes Cavell**. I was given a flash company car, money, security ... frustration, hell. Luke Spikes wanted us to re-invent abstracting! And then Tony died. It was bad times. I spent a period using content analysis to study and publish reports on CRM. It was boring and not very profitable, but it kept the wolf from the door, and I learned very early about community and buyer controlled of marketplaces and their implications which are behind much of the Web 2.0 phenomenon.
- Things began to improve when I wrote the very profitable **‘Taxonomies: Frameworks for Corporate Knowledge’** with David Skyrme in 2001. I am in the process of researching the Third Edition.



- So about 10 years ago we, concluded that we were ahead of our time. Our innovation was to apply the content analysis methodology as a group study tool ongoing as a periodical ... but the people and technology costs were far too high to run a global “Strategic Content Analysis Network” or “Information Refinery” without massive investment. The first Web bubble burst. It was time to retreat to Devon. And as they do, times changed. We now believe that with the advent of Web 2.0 the time is right.

Web 2.0 and Content analysis

Vast Resources

- Millions of people keen to network and share e.g. Clipmarks, Wikipedia, YouTube, FaceBook etc
- Industrial strength Open Source Software available e.g. LINUX, PHP: MySQL, RSS
- Untapped wealth of knowledge available

Growing Problems

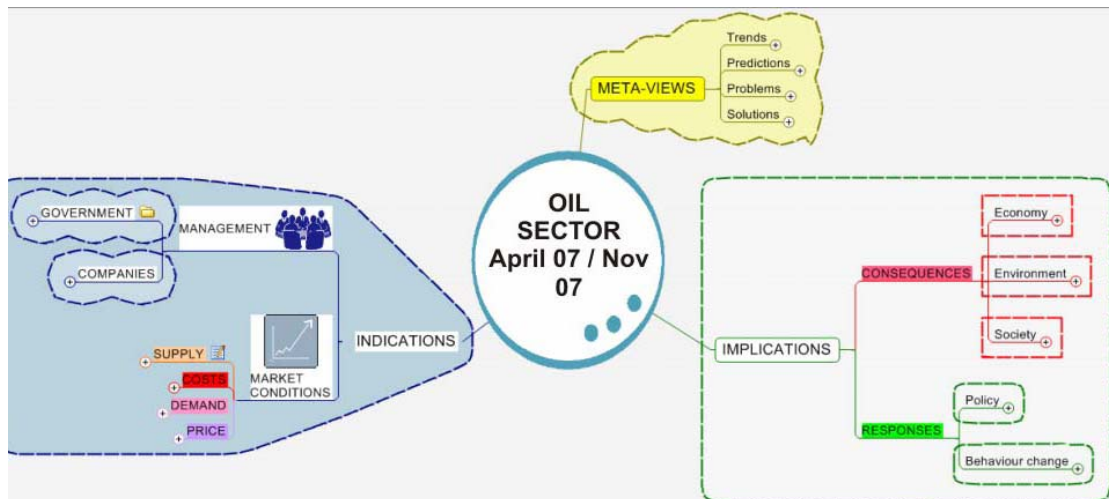
- Increased burden of information overload, more confusion
- Lack of clarity and concentration
- Key knowledge lost, never to be found again

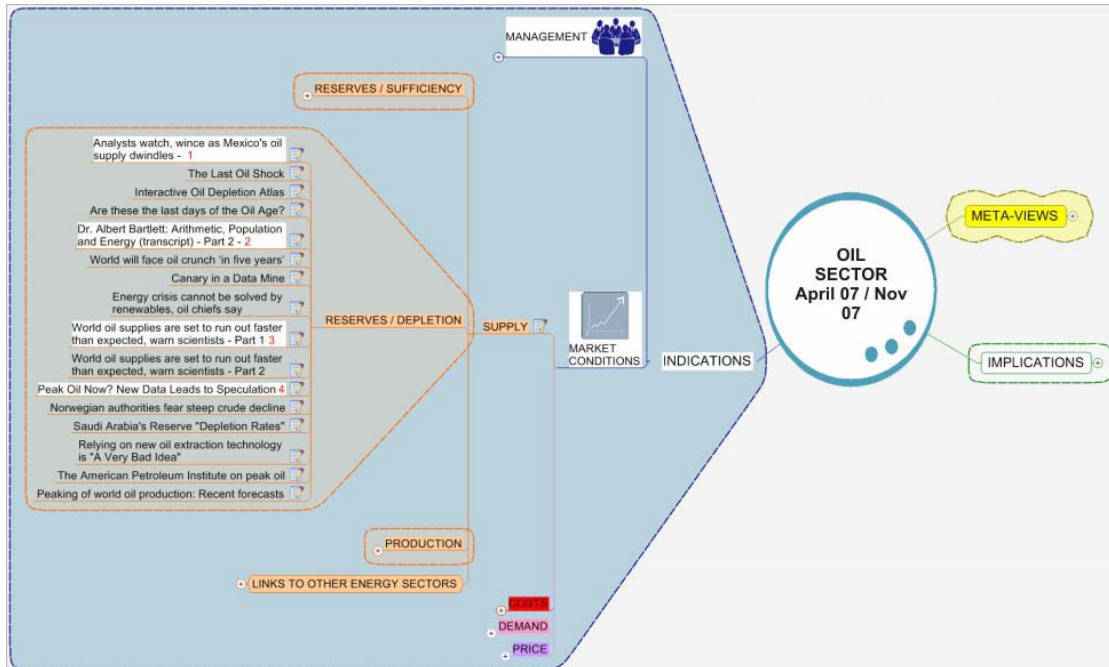
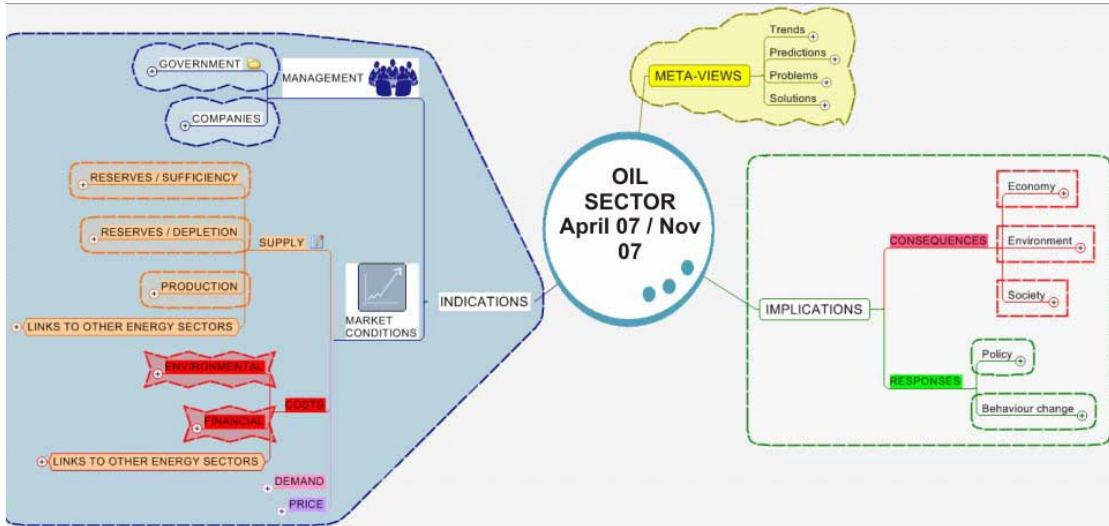
Benefits of content analysis

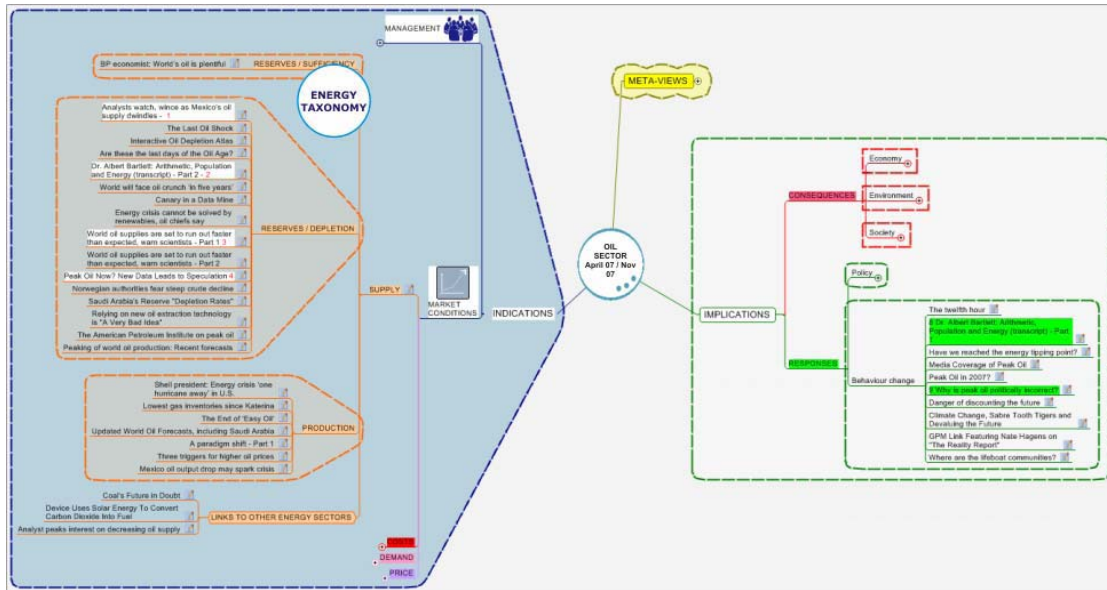
- Thrives on large quantities of information
- Asks multiple questions of any source base
- Channels information flows as they happen
- Finds, highlights and summarises key points
- Identifies emerging trends early
- Provides meta-perspectives
- Makes new cross topic links possible
- Ideally suited to collaborative research projects

The Energy Centre and Open Intelligence

- Although, as Jan said, we concluded about 10 years ago that we were ahead of our time, **we persevered**. Jan kept on cutting and classifying articles from his key sources, monitoring his chosen subjects – energy, the global economy, environment, metaknowledge and the latest web stuff (now Web 2.0). I began learning PHP and MySQL to implement the multifaceted classification and group relationship functionality needed for an online collaborative content analysis community.
- Why we chose to start with the **Energy Centre** was that our own internal content analysis indicated six years ago that this (and environment) would now be the hottest topics. We were right.
- As with all theoretical / qualitative content analysis, we started with a multivariate or faceted information model designed to elicit and keep track of key questions pertaining to the domain. Please note *MindManager* is not a database. It can't do fields or counts which is why we need our own software. Initial research suggested that the material collected divided well into *Indications* of what is happening in an Energy Sector, and their *Implications* for other domains.





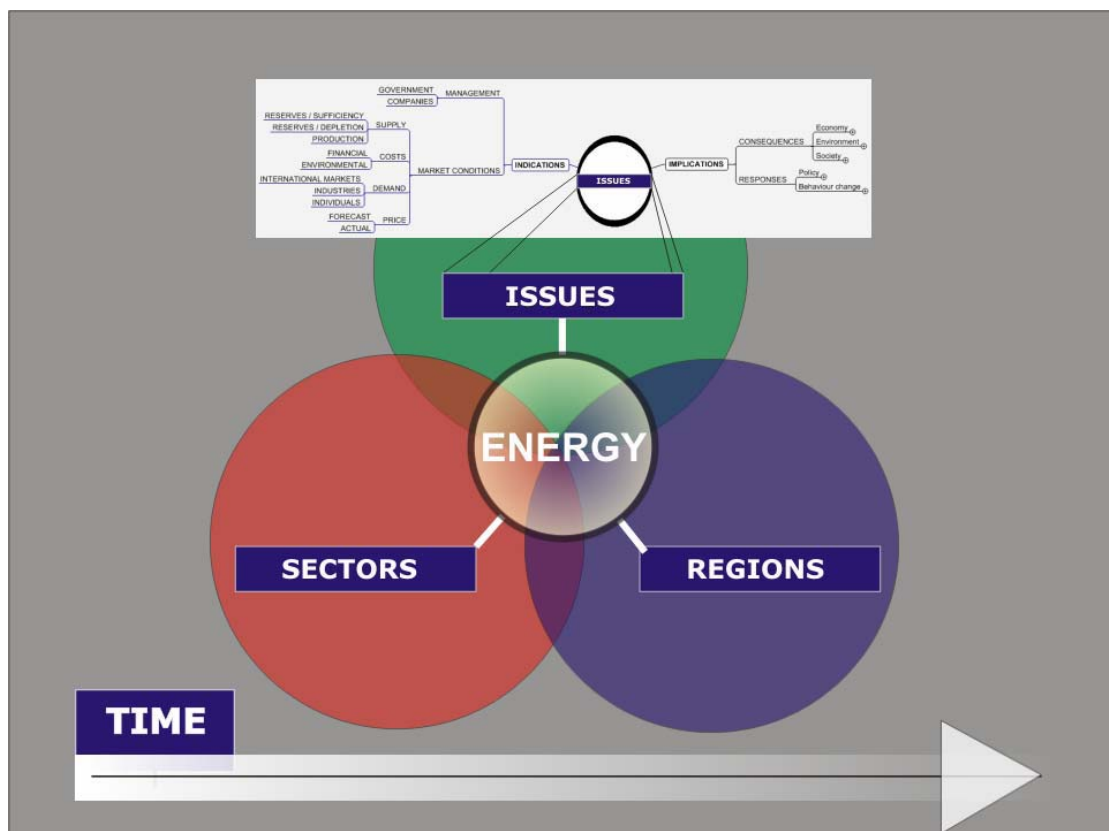


Open Intelligence Software

Software Functions

- Venn simplicity
 - Follow your nose in four dimensions
 - Effortless faceted classification
 - Personal, public and group participation
 - User controlled hierarchical tagging and commentary
 - Tag mapping
 - Automated counting and graphical analysis
 - Multi-level password protected access
- We will now show you how far we have progressed. What we are about to show you is a simulation. The content analysis support software is 80 per cent there, and I predict will be ready for alpha testing early in the New Year.

[A simulated walk through of the software was then given illustrating the above functions, starting with the analysis of the sources using faceted classification and ending with a combination of human inferences and automated graphical outputs. Screen shots are not yet available for publication.]



Many kinds counting and graphical statistics can be envisaged, but just as valuable are the unique concentrations of key quotes on the topics being monitored ... if you want an update of what is being said.

Concentrating the Collaborative Mind

- So this kind of process can be **applied to any subject domain** if you have collected a good enough source sample. In keeping with the times we call the process Open Intelligence ... strapline *concentrating the collaborative mind*.
- In the tradition of Web 2.0, **users create the free content** (as per the very successful Clipmarks). Our Energy site will provide a **meeting place** for groups and individuals to work in an environment of concentrated and clarified information presented in a way to highlight the key questions. We will also give individuals and groups the facility to develop and use their own faceted (hierarchical) tagging schemes.
- So, if in the Web 2.0 tradition, Open Intelligence is free to individuals and voluntary groups, how do we plan to make it financially sustainable? At least, two ways, 1) Google-type **advertising**, and 2) nobody said that passwords would be free for **companies** and **governments** who will be asked to pay for more refined and customised information. Remember in the 1970s, good governments and smart corporations paid thousands of dollars for the much less sophisticated Canadian Trend Report service.