

Beyond Relevance

Patrick Herron

2010 Not Too Far Away

- The Google style of relevant search results with extracted elements has been dominant for at least a decade (precedes Google)
- I expect that to be true in 2010
- My ideas for what I'd like to see rather than what I expect to see

Relevance

- Searches currently typically return relevant results
- Results are regularly portrayed in list form, with hyperlinks, and some extracted element or another from the results

The screenshot shows the NCBI Nucleotide search interface. The search query is 'APOE', and the results are displayed in a list format. The page includes a navigation menu on the left with options like 'About Entrez', 'Entrez Nucleotide', and 'Entrez Tools'. The main content area shows the search results for 'APOE', with a total of 3989 items. The first two results are highlighted:

- 1: [NW_619658](#) Reports Links
Bos taurus chromosome Un genomic contig, NCBI superscaffold, whole genome shotgun sequence
gi|61560651|ref|NW_619658.1|BtUn_WGA454491_1[61560651]
- 2: [NW_526026](#) Reports Links
Bos taurus chromosome Un genomic contig, whole genome shotgun sequence
gi|61466686|ref|NW_526026.1|BtUn_WGA360859_1[61466686]

Beyond Relevance

- Relevance yields too many results
- We need processing of relevant results
- Text mining's job to produce other types of representations – beyond relevance
- Personalized processing of relevant results-
- Visual & integrative elements to take us beyond

Personalization

1. Relevance – for me
2. My preferred representations
3. Custom extraction templates & template builder

Query modification by history & by user preference

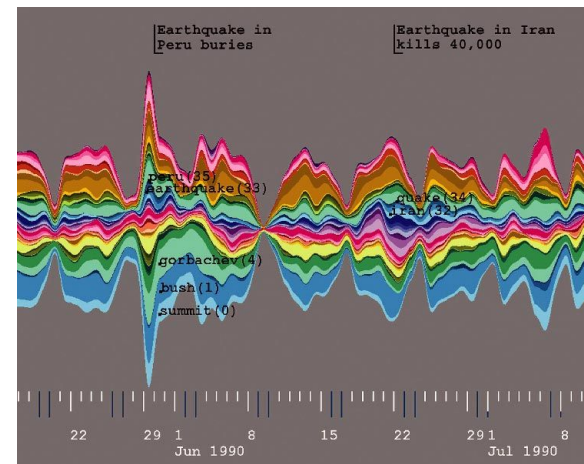
Representation of processing of relevant sets of documents

Visual Elements

- Visual citation analysis (Kartoo)

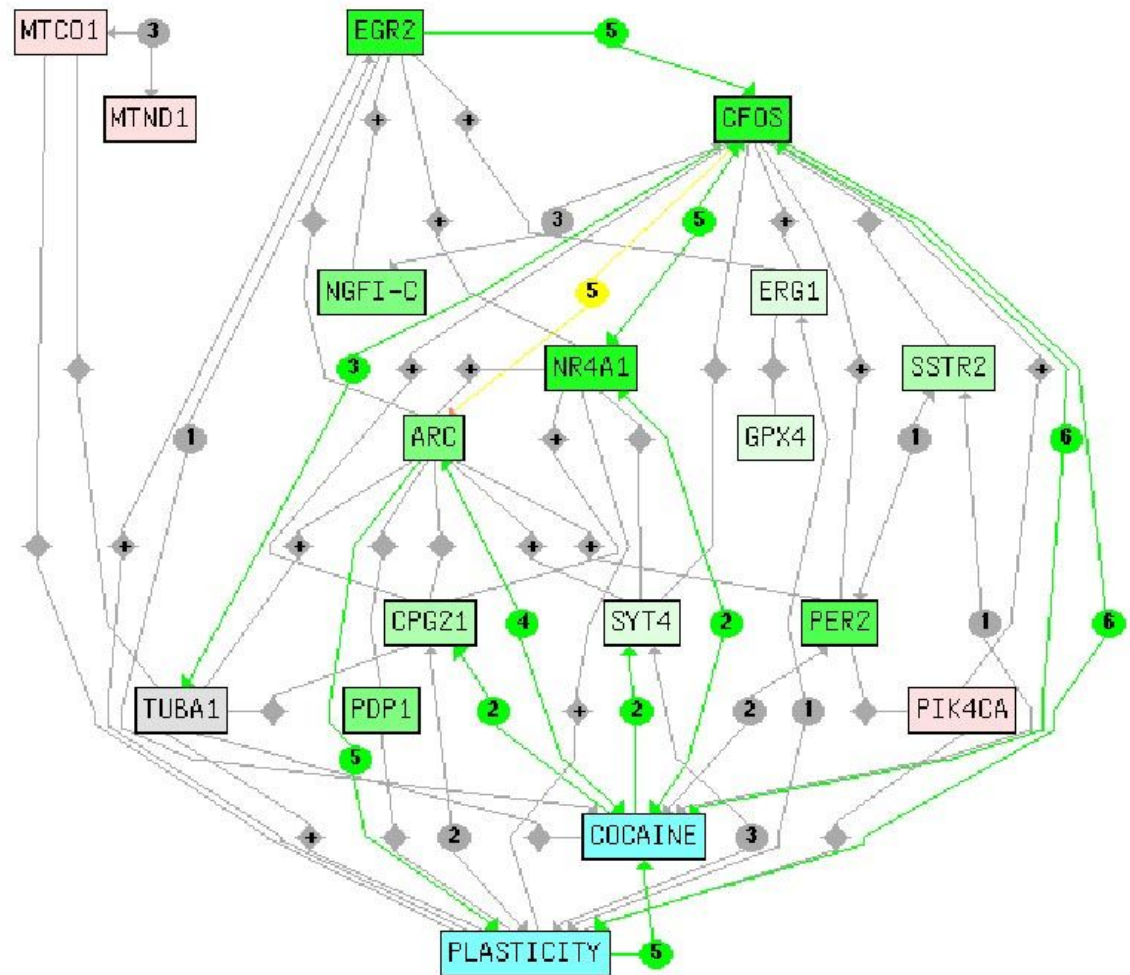


- Visual temporal topic analysis (ThemeRiver)



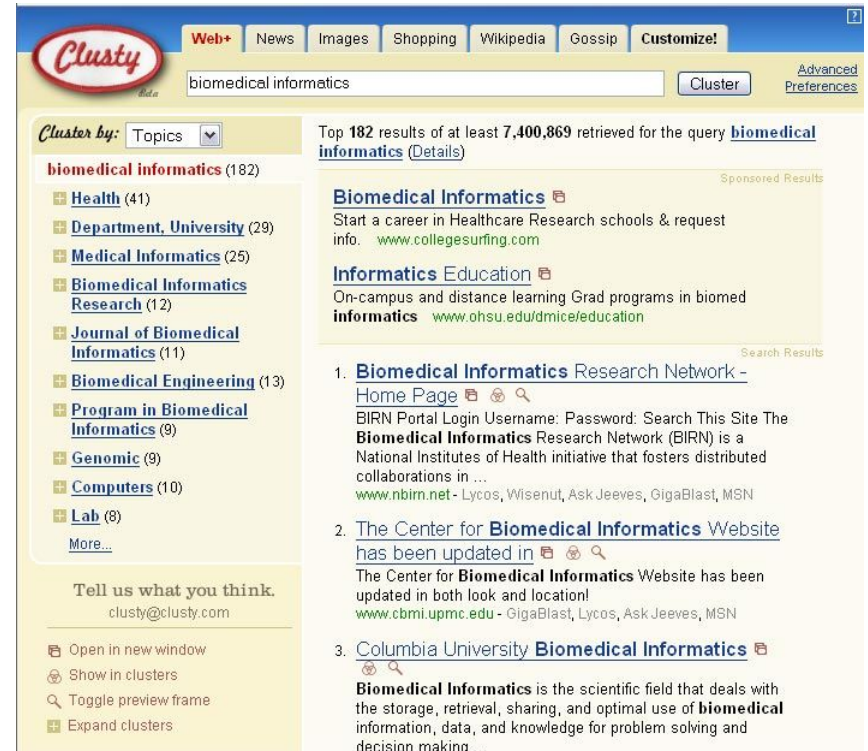
Visual Elements (con't.)

- More topic-based visual network representations (ChiliBot)



Integrative

- In addition to extracted sentences/titles of individual docs, option of brief multidocument summaries based on subsets of relevant results
- Many other text analytic approaches possible – not yet known as to which ones are best
- Clustering-based menus for dynamic IA that organized subsets of relevant results (Clusty); add cluster-based analytics/summaries



The screenshot displays the Clusty search engine interface. At the top, there is a navigation bar with tabs for 'Web+', 'News', 'Images', 'Shopping', 'Wikipedia', 'Gossip', and 'Customize!'. The search bar contains the query 'biomedical informatics' and a 'Cluster' button. Below the search bar, a dropdown menu shows 'Cluster by: Topics'. The main content area is divided into two columns. The left column features a 'biomedical informatics (182)' section with a list of clusters: Health (41), Department, University (29), Medical Informatics (25), Biomedical Informatics Research (12), Journal of Biomedical Informatics (11), Biomedical Engineering (13), Program in Biomedical Informatics (9), Genomic (9), Computers (10), and Lab (8). Below this list is a 'Tell us what you think.' section with the email 'clusty@clusty.com' and four interactive options: 'Open in new window', 'Show in clusters', 'Toggle preview frame', and 'Expand clusters'. The right column shows 'Top 182 results of at least 7,400,869 retrieved for the query biomedical informatics (Details)'. It lists three search results: 1. Biomedical Informatics Research Network - Home Page, 2. The Center for Biomedical Informatics Website has been updated in, and 3. Columbia University Biomedical Informatics. Each result includes a brief description and a URL.

Schizophrenia

- Maybe want to search relationship between 3D brain maps, genes and compounds across chem databases
- Relevant set of documents doesn't exactly make sense
- Associated term scapes, associated gene & protein networks, multidocument summaries

Summary

- **Likely** to see relevance continue to dominate
- **Hopefully** there will be client-side history for purposes of query modification and choice of visual & integrative representations for processing of relevant results (beyond relevance)
- Text mining affords opportunity for integrating subsets of relevant results
- Visual elements hopefully will become more dominant and more useful