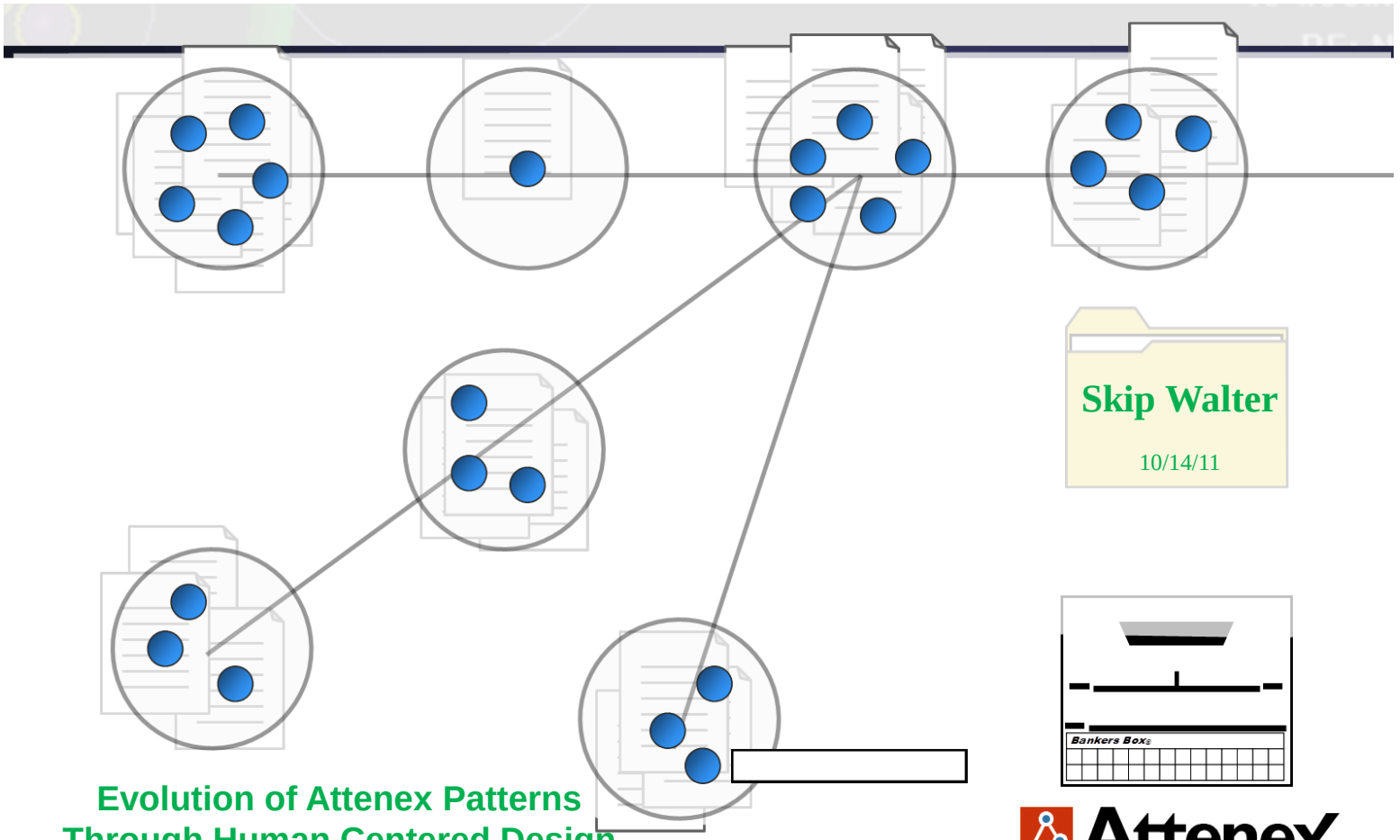
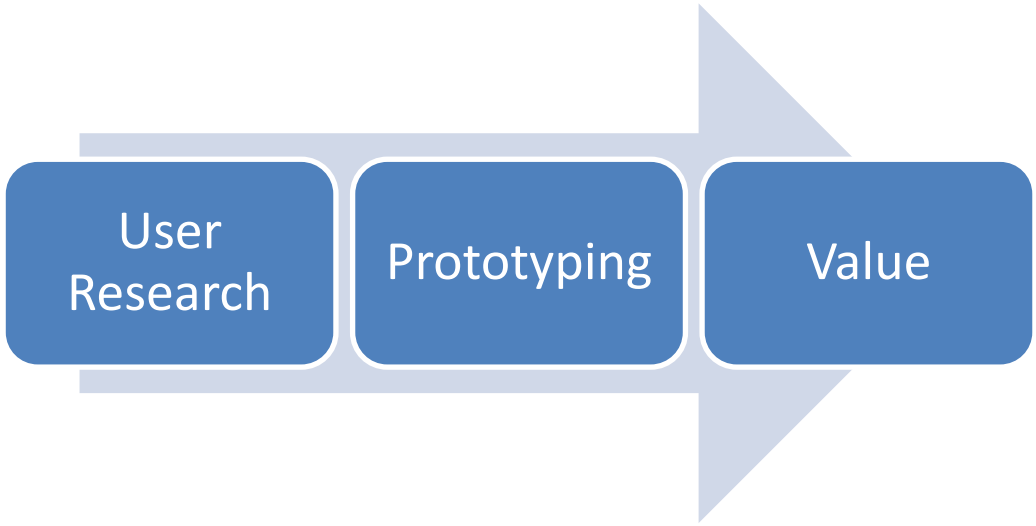
Seeing What Matters



Evolution of Attenex Patterns
Through Human Centered Design

Attenex

This briefing book documents the development and evolution of an extreme productivity application for electronic discovery, Attenex Patterns, in the litigation market. The steps in the process are captured below:



During the last 5 years, this briefing was given in conference settings, academic settings (seminars and classes), consulting presentations, and entrepreneur mentoring. The feedback from designers, product developers, executives, professors, and students consistently echoes that this is the first time they've ever seen a presentation on the making of a product to the "go to market" plan to the financial results and ultimately to the valuation of Attenex and its products.

Attenex Patterns is proof that visual analytics can lead to dramatic increases in productivity.

The purpose of this briefing book is to show the entire process of extreme productivity in product development and to provide insights into the power of visual analytics.

Customer Context
Creating a Legal eDiscovery Tool

- Your customer(s) are law firms providing legal review of paper and electronic documents for Fortune 1000 clients
- Your task is to identify “responsive” and “privileged” documents
- Your problem is that the current method of review (paper) is too slow and too costly

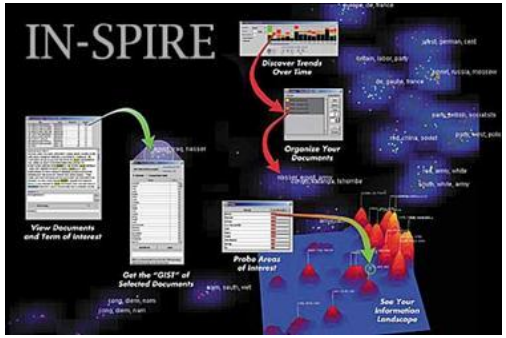
HCDE Human Centered Design & Engineering

User Research

The first part of user research is working with the client to get a clear idea of the problem statement and what the research task is.

In 2000, Preston Gates and Ellis (now K&L Gates) was the primary electronic discovery review vendor for Microsoft. As an early user of its own products, Microsoft generated hundreds of gigabytes every year of electronic material that was potentially discoverable in its DOJ anti-trust litigation along with several other lawsuits. Microsoft was seeing its cost for eDiscovery escalate each year. So William Neukom, Microsoft General Counsel, asked Preston Gates to come up with a cost reduction strategy either by using less expensive review resources, moving the review off shore to India or finding a technology solution.

Preston Gates decided to pursue a technology solution based on research they became aware of at Pacific Northwest National Laboratory in Richland, WA – IN-SPIRE (formerly SPIRE).



Prior to pursuing a technical solution, Preston Gates had moved from a pure paper review of documents to an online review of documents one page at a time. Both methods required the same amount of reviewer time, but with the online method Microsoft didn't have to first print every document out.

So the challenge that Preston Gates took on was to see if there was a way to improve eDiscovery through visual analytics.

The following video illustrates the labor intensiveness of legal document review.

FACTOR10x

Paper Based Review

User Research

The next step in the user research process is to observe the target users in their task environment. This kind of user research can range from pure observation to video ethnography to participatory research. See Chapter 13 (Field Studies) of Courage and Baxter's *Understanding Your User* for a description of the types of user research.

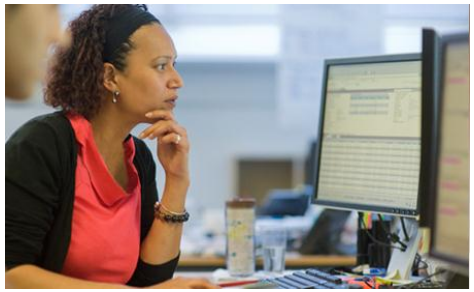
The first part of the video shows a warehouse full of bankers boxes which can hold up to 2000 pages of paper. The supervisory attorney selects the next box of documents to review and then passes them on to the reviewing attorney.

The reviewing attorney then pulls out a folder of documents. Along with "yellow stickies" to mark the category of a document (responsive, non-responsive, privileged, or hot) the attorney has the case intelligence notebook which provides criteria and examples for categorizing a particular document.

When the attorney finds a responsive document, he then goes to the copier to make a copy of the document to put into the Responsive folder.

And the process repeats.

Personna: Sandra - Review Attorney
Dewey, Cheatem, and Howe, LLP



- Top Requests
  - Easy to access
  - Easy to use
  - Trustworthy
- Goals
  - To perform well and build a good reputation at the firm.
  - To learn the ropes and help out where needed.
  - To begin to establish a successful practice of her own.
  - To pay off law school.
- Job Description
  - Sandra is a second year associate attorney at Dewey, Cheatem and Howe. The firm is based in San Francisco, but Sandra works in Seattle with nearly 60 other attorneys.
  - She spends much of her time writing relatively straightforward contracts or assisting more senior attorneys in writing portions of complex contracts.
  - When large litigation cases come into the firm, Sandra is occasionally summoned to help out on an "all hands on deck" discovery effort. Often, the case is based in another office and she is one of many attorneys scattered in various places "pitching in" to complete the review process.
- About Sandra
  - Sandra is a young lawyer, very dedicated to her career, happy to finally be practicing law after a long time in school.
  - Sandra is relatively comfortable with MS Office applications. She got her first e-mail account her senior year in college. Law school definitely familiarized her with the web. When something goes wrong with her computer, however, she calls in the IT department for help.
  - As a new associate, Sandra has very little free time but she does get in an evening of ballroom dancing every once in a while.
  - Age: 29
  - Car: Seattle Light Rail

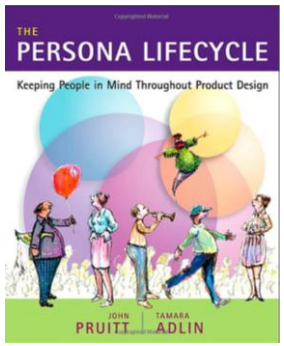
5

User Research

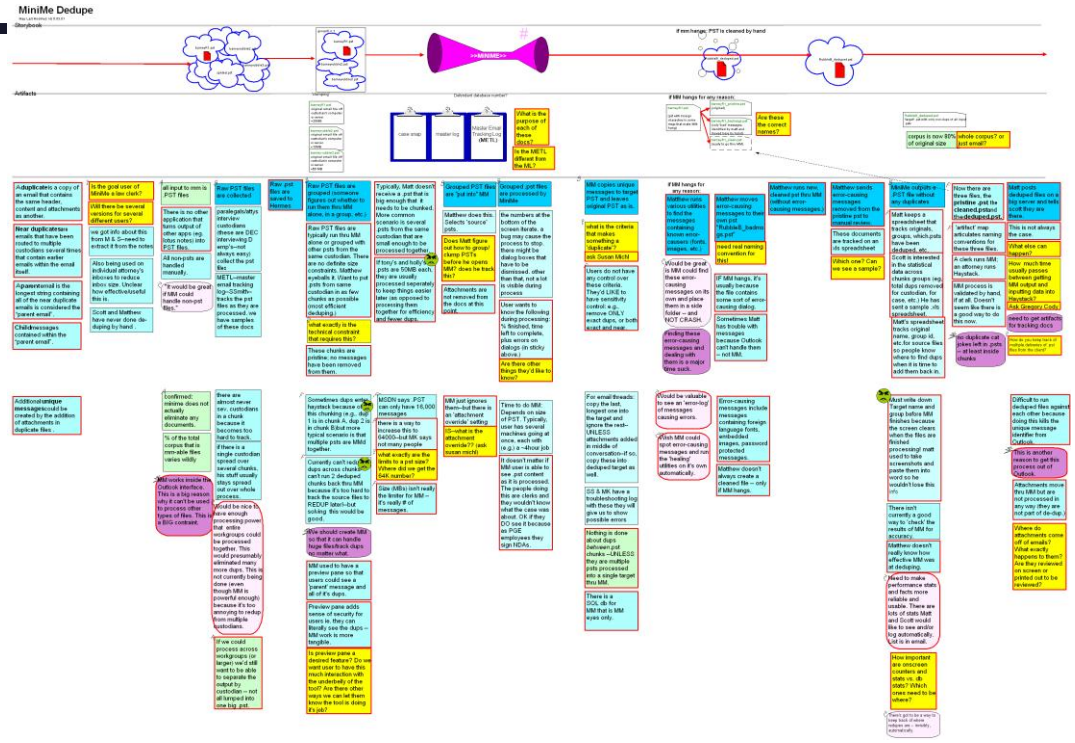
The next step in the user research is to build personas of the different types of users.

The above example focuses on the reviewing attorney as the key persona in the early user research. While there are multiple personas in eDiscovery (the project lead, the litigating partner, the data wrangler), the reviewing attorney persona has the most users and is far and away the bulk of the billable hours.

An excellent reference on personas is: *The Persona Lifecycle* by John Pruitt and Tamara Adlin (http://www.amazon.com/Persona-Lifecycle-Throughout-Interactive-Technologies/dp/0125662513/ref=sr_1_1?s=books&ie=UTF8&qid=1318523084&sr=1-1).



First Map the Process

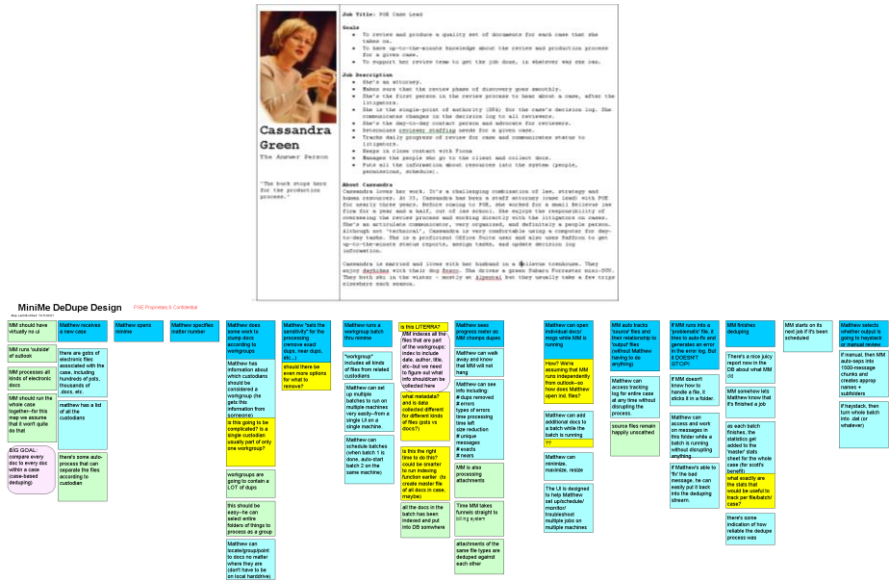


User Research

The next step in the user research process is to create a map of the work processes and task processes.

Multi-colored post it notes are a great way to map the current state of the process. One color is picked to show what the key "bottlenecks" are in the workflow. This research step builds on the work of Eli Goldratt whose Theory of Constraint illustrates that there are only 2-3 bottlenecks in any complex workflow. Instead of trying to solve for and optimize each step in the workflow, Goldratt shows that you only have to optimize and regulate the 2-3 bottlenecks.

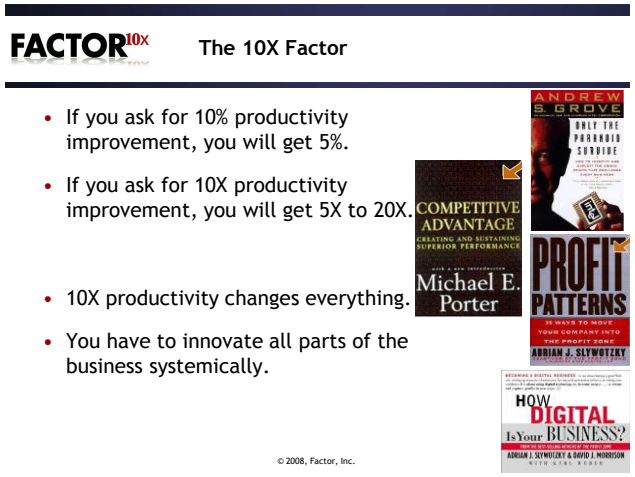
…now, re-write the story. Increase Productivity



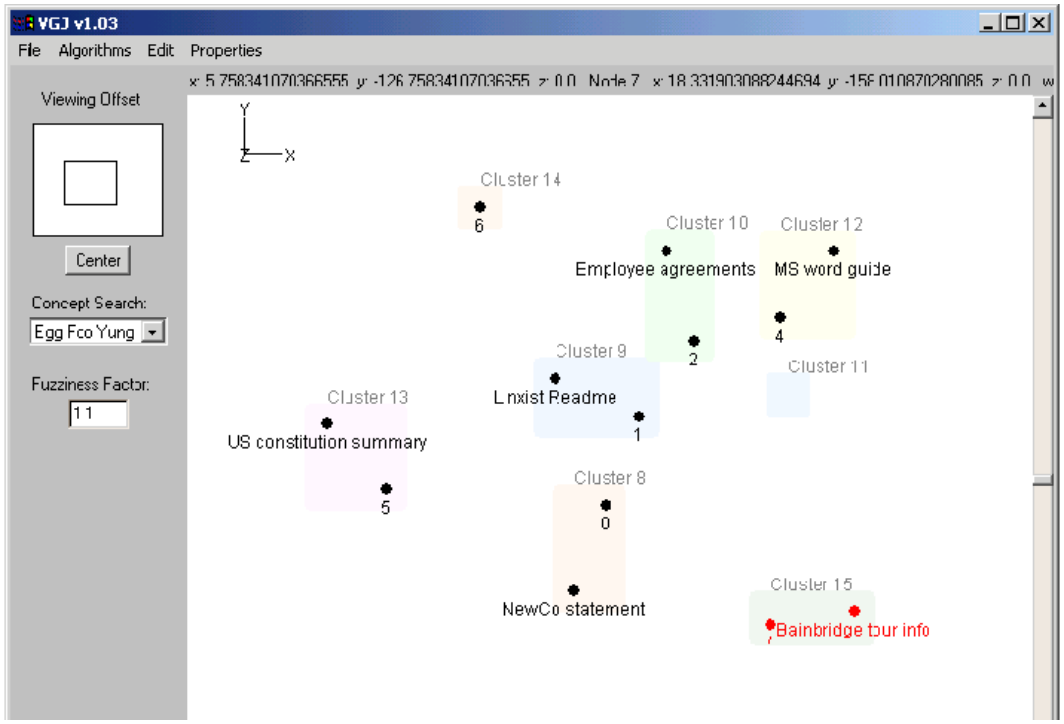
User Research

After you capture the today state of the workflow, your then create an Idealized Design Desired State for the new workflow that focuses on improving the bottlenecks. A good redesign of the work process should eliminate all the Purple Constraints from the Today State.

Unlike all of our competitors in the eDiscovery technology and tools market, Attenex focused on really improving (at least 10 times productivity) the few steps in the process that mattered. The rest of our competitors tried to create functionality for each step in the process with the result that they only marginally improved overall productivity.



6/1/01: Very First Interface
Organize and Cluster Documents in 3D



Prototyping

Once the human centered design team has gotten a sufficient set of user research accomplished, it is time to start the Rapid Prototyping phase.

As good technologists, we felt that a 3D interface was the right way to go. The PNNL InSpire product was a 3D visualization and always got great responses when it was demoed. However, we quickly found several problems with trying to generate a 3D visualization. The first was technical – there just wasn't enough computing horsepower on the desktop in 2000 to quickly visualize even small document sets in 3D. The second part was human – our target reviewers were lawyers and they are words people. As we had the reviewers work with the 3D interfaces they were constantly getting lost in the three dimensions and couldn't figure out how to navigate at all.

We tried for the longest time to make a 3D interface work but productivity always suffered.

The good news that came out of this process was that we came up with a clear productivity metric – document decisions per hour. Going forward every time we added new functionality we would immediately test with our target users whether the new functionality increased or decreased document decisions per hour. If the metric improved, we left the functionality in. If productivity went down, we took the functionality out.

The dimensions of productivity we identified are:

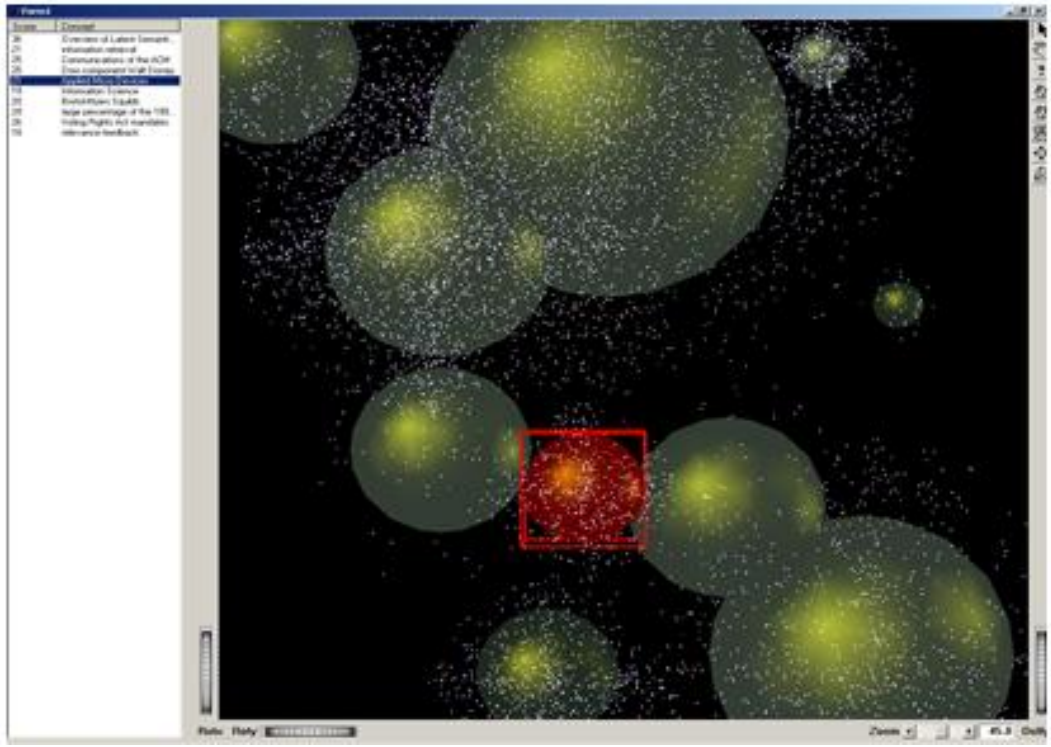
- Faster
- Better
- Cheaper
- Increased Quality
- Improved relationship capital
- Risk Management
- Increased reward

Faster has several dimensions:

- Cycle time – the overall number of days, weeks or months it takes to complete a project from beginning to end
- Process time – the total number of hours actually consumed in producing the end work product
- Learning time – the amount of time required to achieve proficiency in a particular area.

FACTOR10x

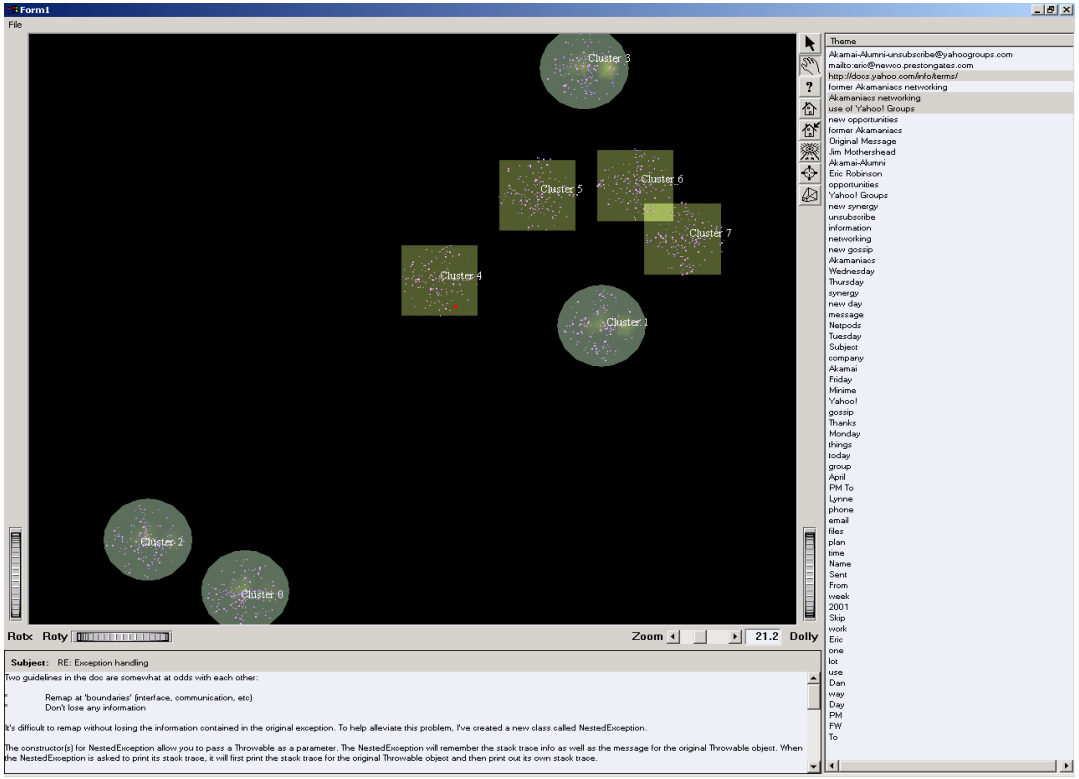
6/7/01: SeeMore Sketches



Prototyping

The next prototype while still in a 3D framework improved the content analytics so that we could now have 100s to 1000s of documents displayed (each document is a single dot in the display above). The spheres represent clusters of documents with the red colored sphere being the selected set of documents. On the left hand side of the display are the key concepts which caused the documents to cluster. The user selected "Applied Micro Devices" and that is the sphere lit in red.

6/19/01: The First Real Clusters!



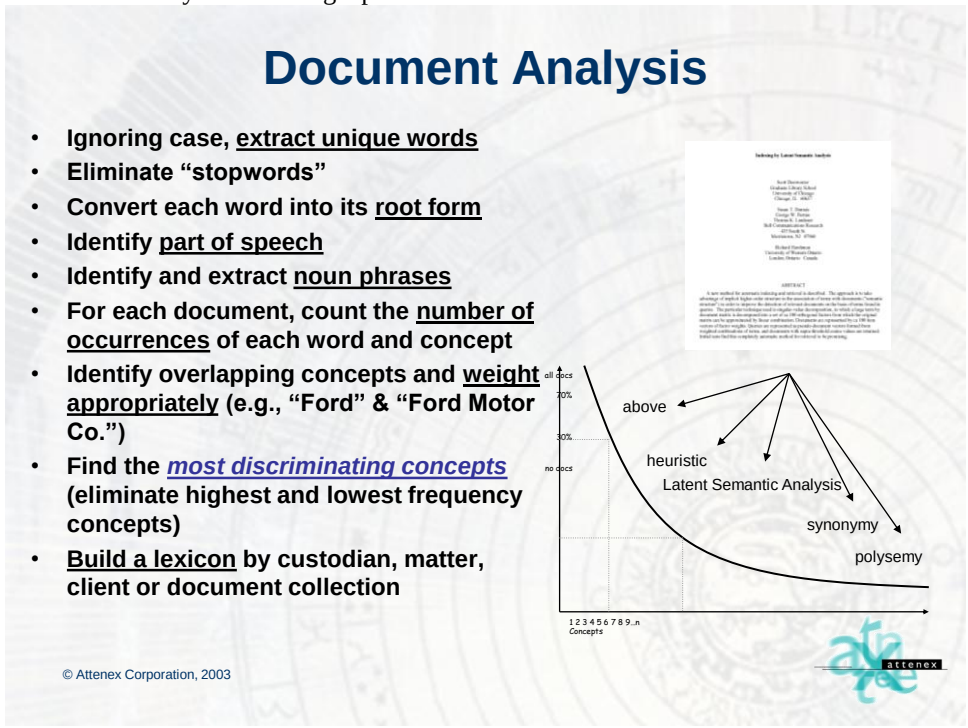
Prototyping

While this prototype still uses a 3D interface (build on OpenGL), this display starts the process of focusing more in a Tufte sense of displaying multiple variables in a two dimensional space.

We start to add a couple of windows in this display. At the bottom is a window displaying the text of a selected document.

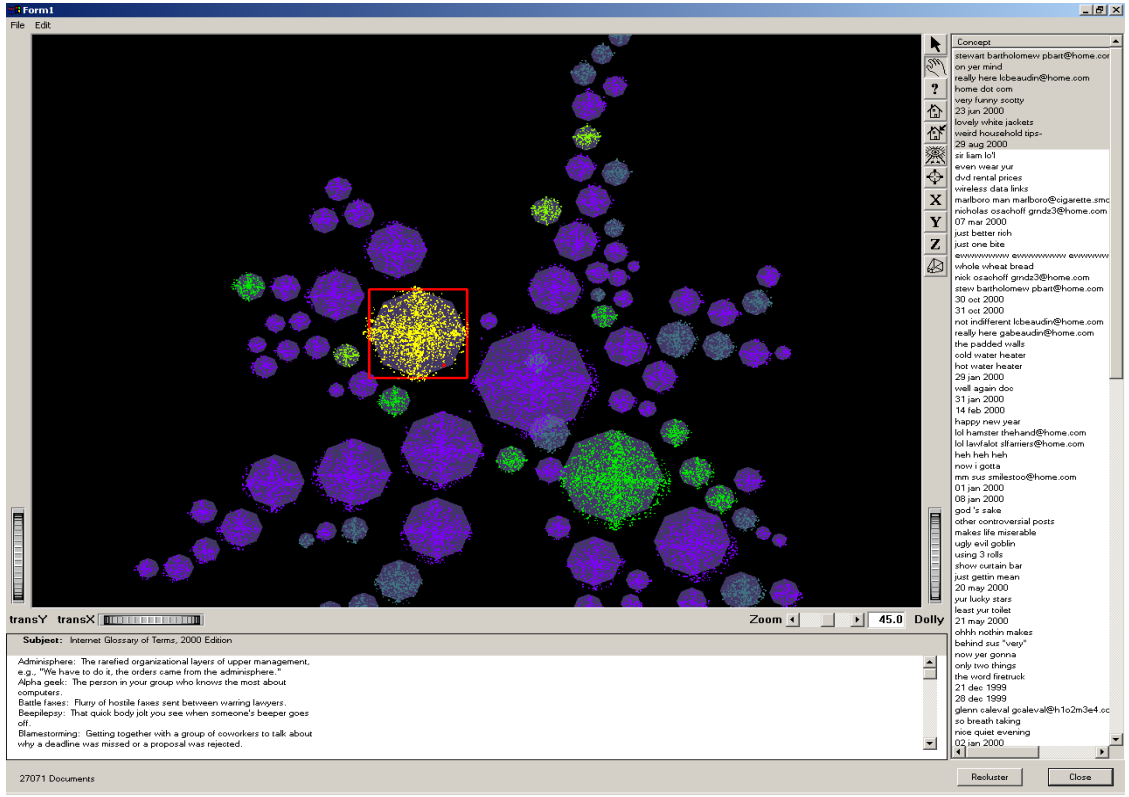
On the right hand side of the display the concepts that are pulled from the documents are displayed on the right.

The content analytics are using a process described below:



For legal applications we found that part of speech tagging was more efficient as lawyers typically organize things by nouns or noun phrases. So that's what we clustered on. For other applications like resume processing it turns out that the verbs are most important, so one would throw away the nouns and only keep the verbs for clustering purposes. For applications like customer support, both verbs and nouns are important.

7/13/01: Proximity Clustering



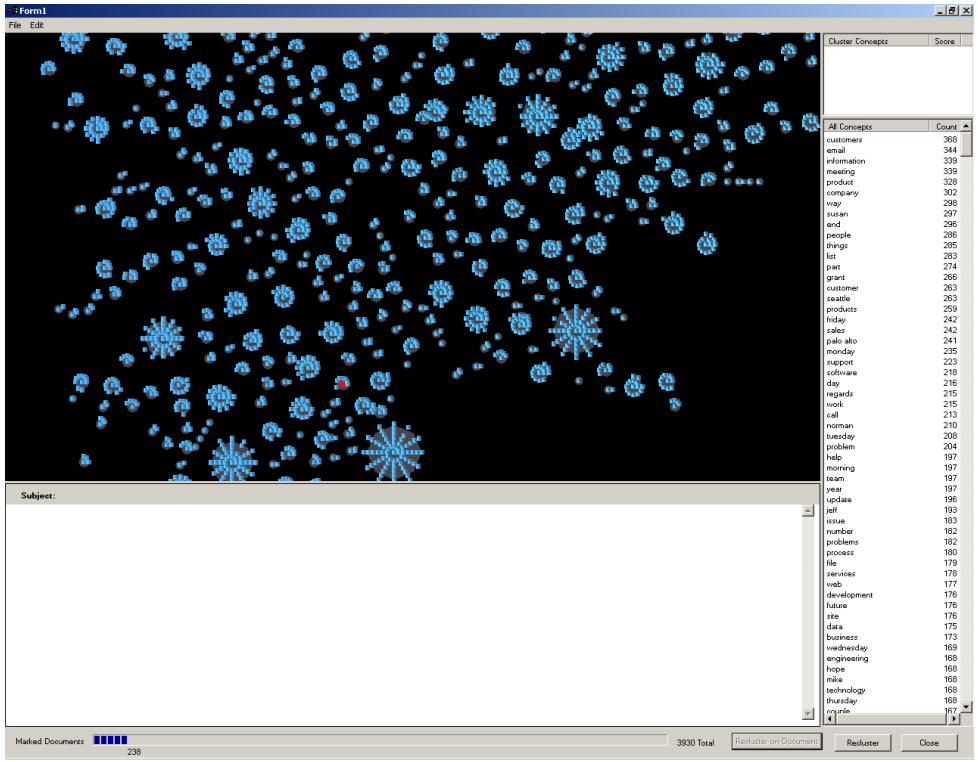
Prototyping

In this prototype, we moved to using proximity of key concepts both for organizing documents within a cluster (the dots) and between clusters.

We found that proximity worked well for organizing how clusters were layed into the two dimensional space, but did not work well for displaying the document "dots" within a cluster.

In the concept window on the right, we started segmenting the concept list into the key concepts that are within a cluster (the top part of the list) and then the concepts that occur across all of the documents.

8/6/01: Orient in clusters



Prototyping

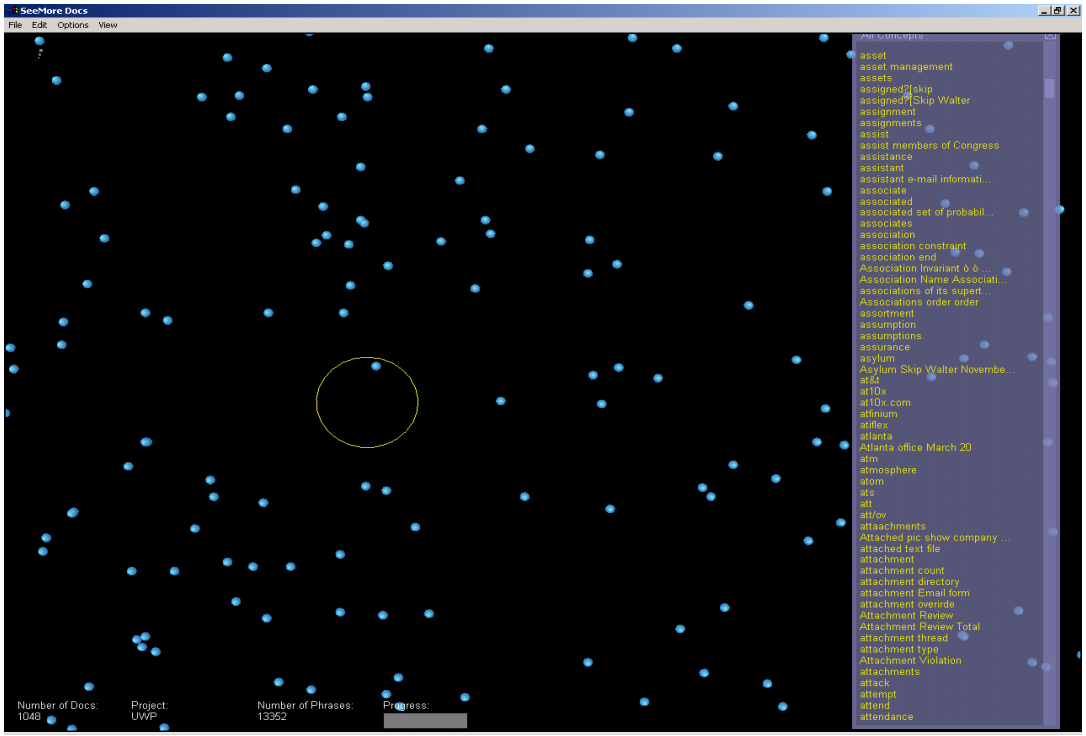
This set of prototypes looks at different ways to orient the documents within a cluster. For the purposes of the prototype we just did random snowflake patterns of documents. We then used this prototype as a kind of paper prototype and tested a number of users to see if one or more of the ways to organize within a cluster could make sense.

We did a think aloud protocol and asked the users what they thought the organizing principle was for each arrangement.

Consistently, the users found that the spiral orientation made the most sense to them:

Unaided the users consistently said that they would expect that the document (dot) in the middle would be the most on point to what caused the documents to cluster and the one at the outer edge would be the least on point. Thus, to see what category the documents might be all you would have to do is check the document in the middle for a category (like non-responsive) and then the dot at the outer edge. If they are both non-responsive then you know that the whole cluster is non-responsive.

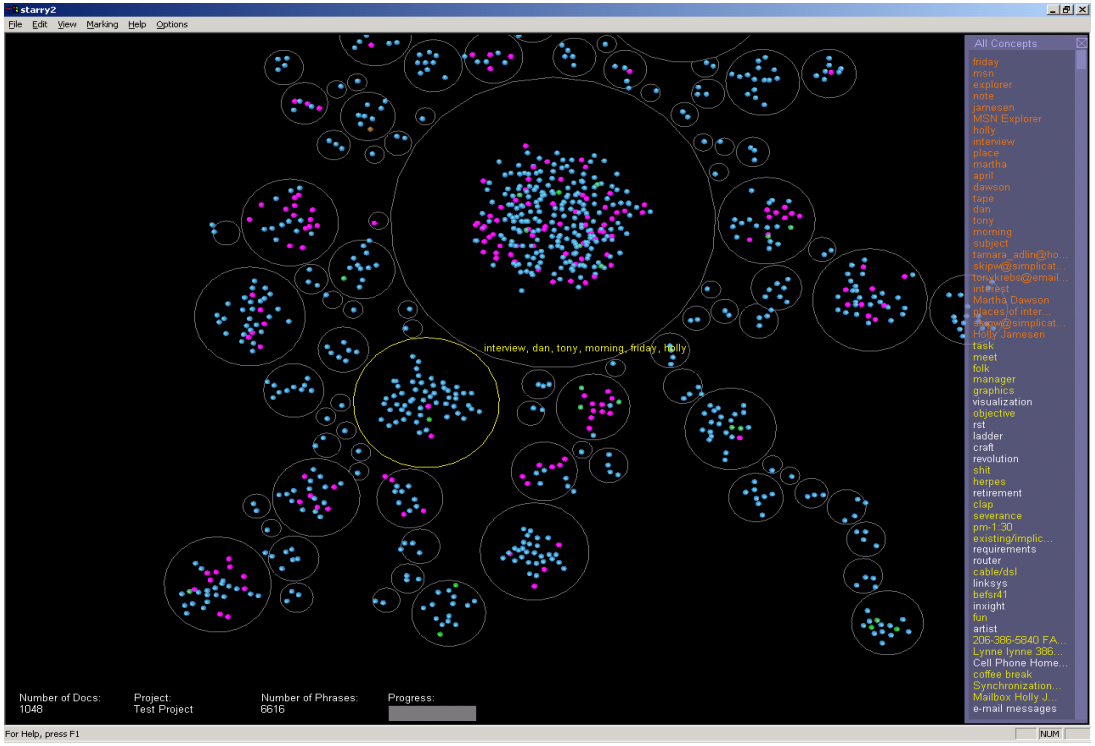
9/14/01: Transparency!



Prototyping

The transparency prototype experimented with making the concept pain translucent so that more of the screen real estate could show the document clusters through. In 2001 when we were working with these prototypes 1200x1000 screens were the standard (higher resolution was very expensive). So the more document map screen real estate we could get the less the user would have to zoom in and out.

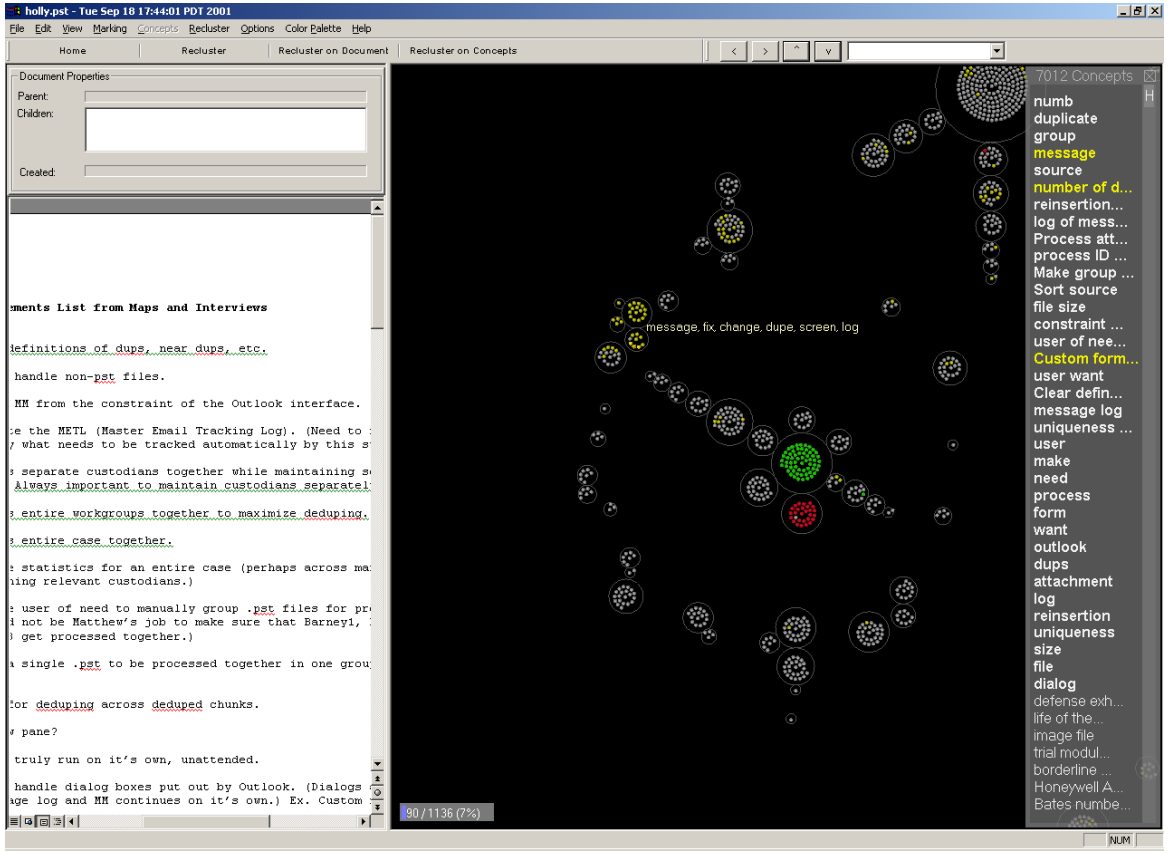
9/21/01: A Little Color Work?



Prototyping

Once we had the basics of how to visualize a document space map, it was time to work with colors. This prototype starts associating colors with the categories of documents and also with what colors should be used to select a collection of documents (filtering through the concepts in the right hand panel).

10/2/01: Docviewer!



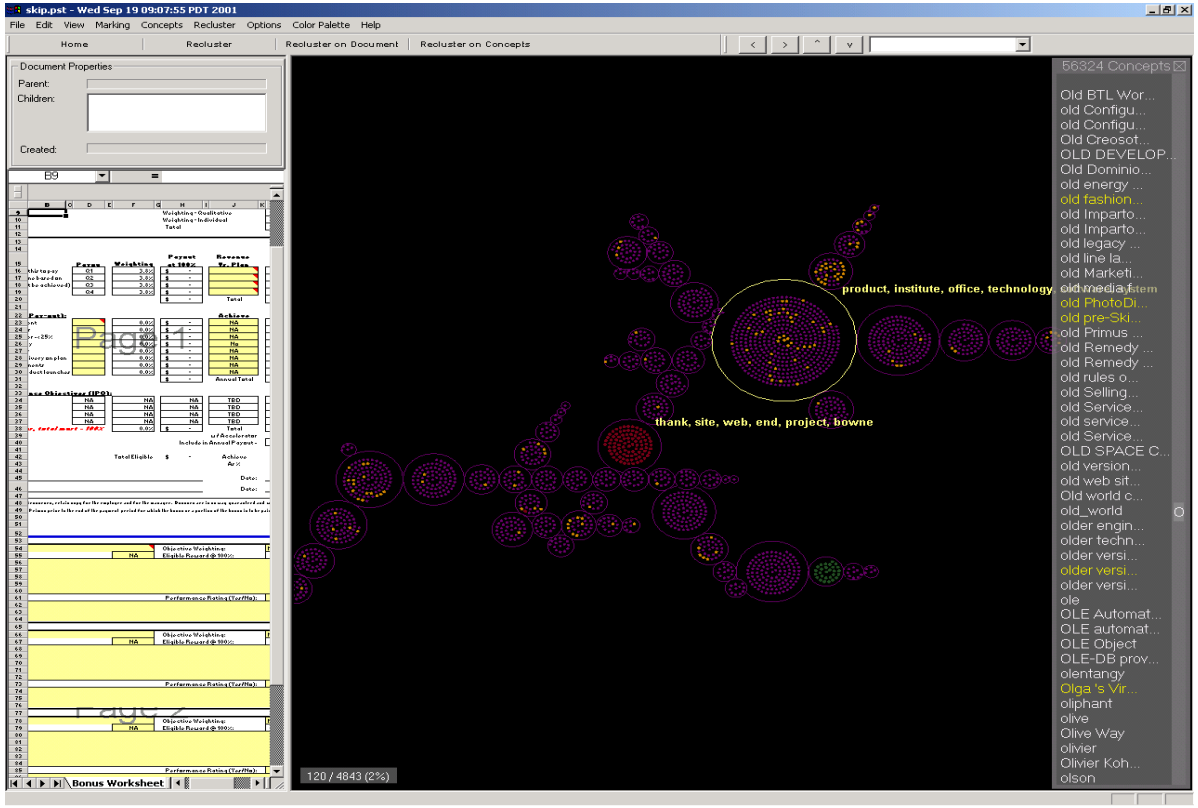
Prototyping

Now that we had the concept map (center pane) in the visualization coming along – spiral for organizing within a cluster and proximity for organizing clusters it was time to focus on what information we wanted to display in the document viewer (left pane).

The document viewer in this first example is a simple display of the text with word wrap in the document with the ability to scroll. In the top left pane you can see the notion of parent and children, that is, what are the relationships of a document within a email thread (message body or an attachment) or within a folder structure.

You can also see the start of the concepts displaying next to a cluster as a cluster is selected or rolled over (center pane).

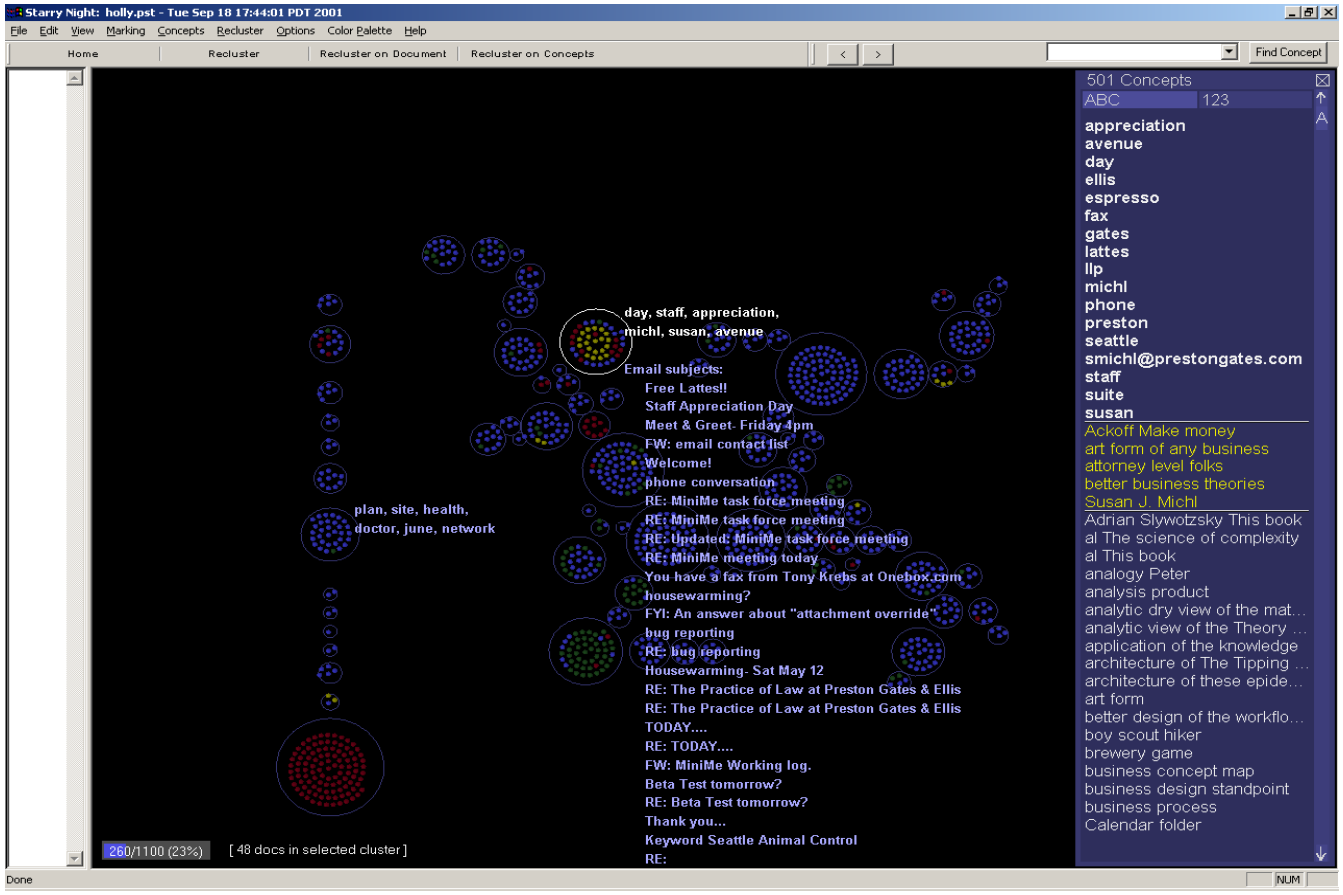
10/3/01: Metadata Organizing Clusters



Prototyping

Within the center pane in this prototype you can see the proximity function of between clusters being organized in spines. There are now two levels of the metadata (concepts in yellow associated with a cluster). The concepts are the primary ones that caused the documents within a cluster to come together. Then one or more of the concepts are used to "spine" collections of clusters.

10/23/01: Email Subjects



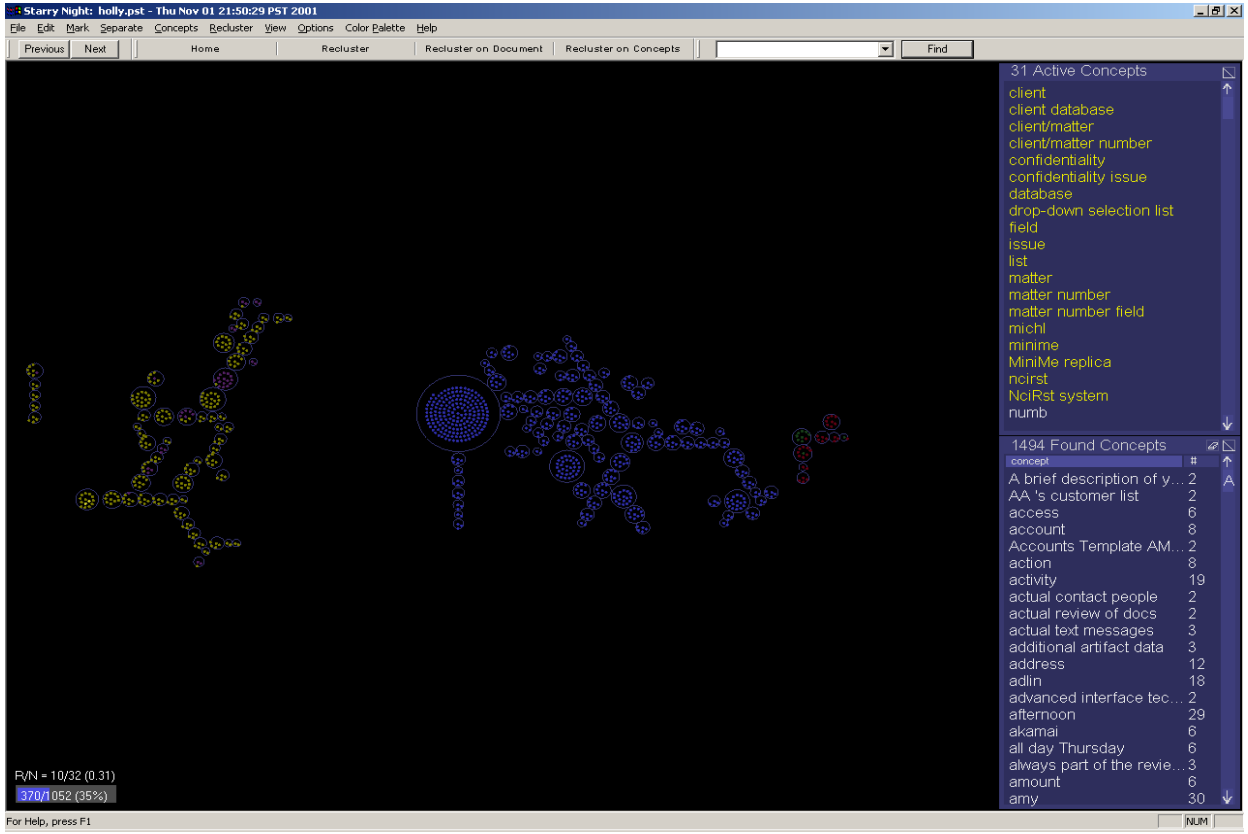
Prototyping

While having the concepts (in bright white above in the center pane) were useful in organizing a cluster, they weren't enough for the reviewer to figure out what a given cluster was about. So we added the email subject line or the full file name of all the documents in a cluster. With the combination of cluster concepts and with the email subject lines a reviewer could quickly figure out whether a given cluster was likely to be responsive to a request for production or non-responsive.

While we had hoped that the visual layout would be enough for the reviewer without them having to read a lot of text, we found that we had to balance the visual map with an overlay of relevant summarizing text.

We also experimented with automated document summarization, but none of the techniques we evaluated did a good job of summarizing medium to long documents and they also generated too much text to look at. We found that reviewers if they needed additional information would prefer to open up the entire document in the docviewer.

11/9/01: Separate Out Marked Documents



Prototyping

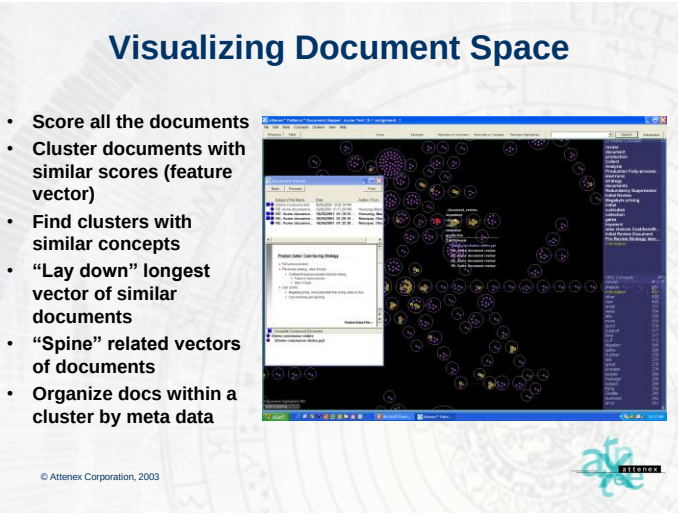
Now that we had the basics of how we wanted the document map to work along with the highlighted concepts and the complete concept list (right hand panes), it was time in the visual analytics sense to work on what to do with the decisions and actions that would be occurring in the workflow.

Once the reviewer figured out what category a document should be (responsive, non-responsive, privileged etc), they would right click on the document or the cluster or the spine and select the marking category. Pretty soon the document map would be cluttered with marked and un-marked documents. So we started adding different ways to recluster and organize the document map. If you look at the second row of buttons at the top of the display you will see multiple ways to reorganize the data (home – cluster all of the documents together, Recluster to separate the marked from the unmarked [what you see above], recluster on the concepts in one particular document, and recluster on highlighted concepts. With each of these ways of clustering, the reviewer could strategically organize the documents to get a better idea of what was related to what based on the quality of the documents that they were finding.

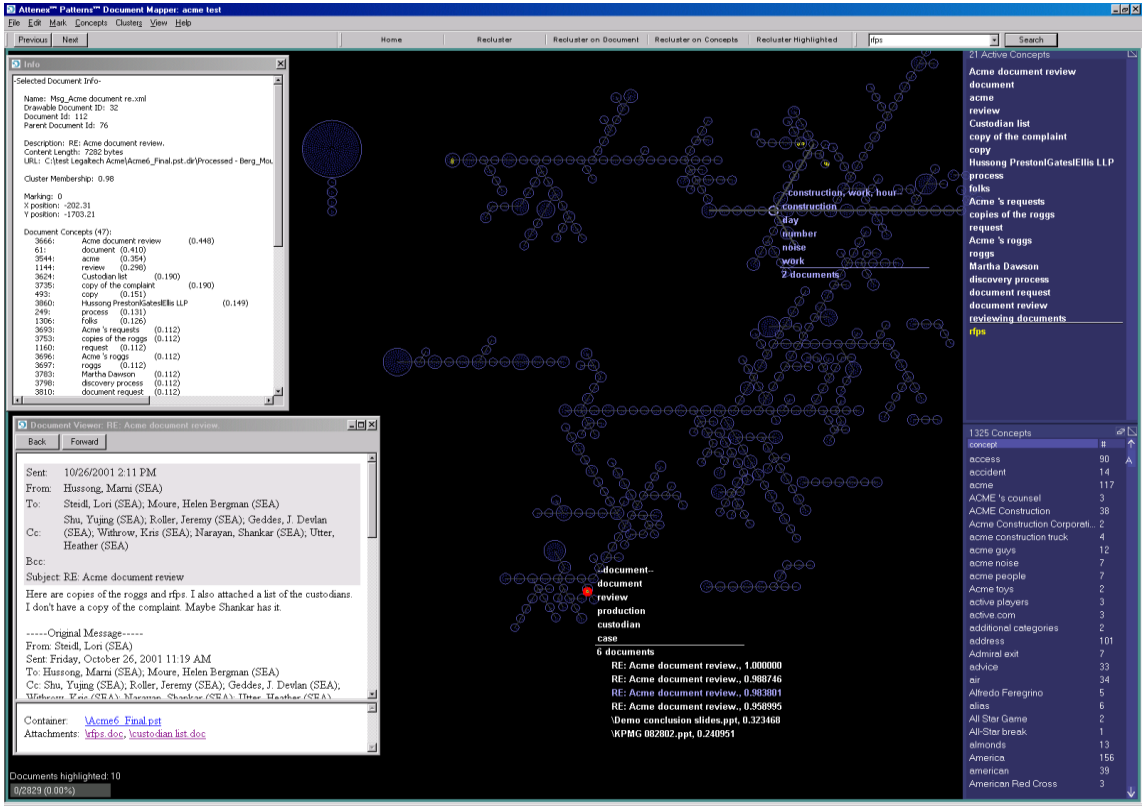
The version represented by this display is what we put into production at Preston Gates and Ellis in November of 2001. We had put our first software in production in May of 2001 (company was founded in April of 2001). The first software was a deduping tool which removed all duplicate files and emails from what the reviewers needed to see eliminated 30-50% of the documents which needed to be looked at individually.

This prototype became the V1 production version.

We then set as a goal to review 4 million documents in at least 10 separate matters before we would publicly announce the product. The following slide looks at the steps to visualize the concept map.



03/03/03: Patterns 1.5



Prototyping

Finally in June of 2002, we released the first public version of Attenex Patterns. The user interface for Attenex Patterns Document Mapper hadn't changed a lot between the internal production release and the 1.5 release in March of 2003. However, the backend part of Attenex Patterns which was called Workbench occupied most of the development team for the over two years.

When we first started the project a large matter was 50Gbytes. However as more and more corporations became depended on Microsoft Outlook/Exchange the volume of documents kept doubling every 3-6 months. We had to continue re-writing Workbench to improve performance to keep up with the ever increasing matter sizes. By the time that V1.5 was released, the average matter size had grown to 500Gbytes.

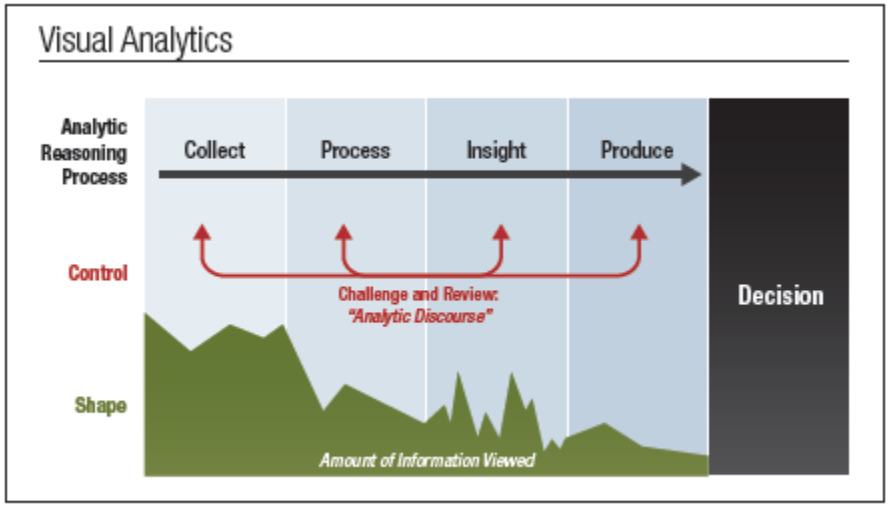


Figure 3: Visual analytics, shape and control chart

Action Research
Developers as Reviewers

23

User Research

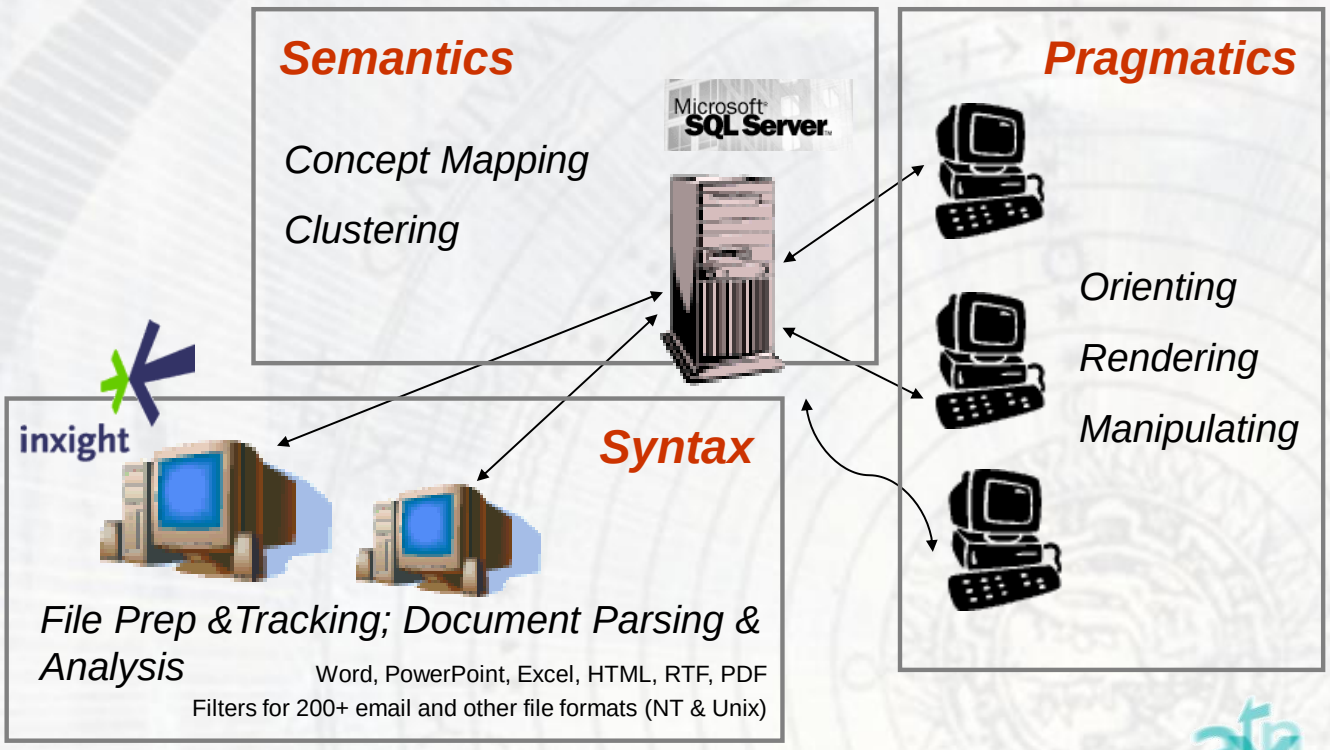
While most of our prototyping was tested with actual lawyer reviewers, periodically we would have the developers spend a day doing participatory (action) research, where they would work with document sets that had already been reviewed by the attorney reviewers.

This photo shows the development and QA team spending a day reviewing documents.

Over the years I've found that the developers can get very out of touch with how the users actually use the product. Video ethnography is an excellent way to show developers how users employ the product. However, there is a rich ideation result that comes from developers using a product on real data. Invariably after one of these sessions, the developers would come up with new productivity ideas that had escaped us.

From a usability and user experience viewpoint, these sessions always helped fix the lingering usability problems. About every half hour one of the developers would exclaim "I can't believe the reviewers put up with this hard to use feature." Magically, the next day the usability issue would be fixed. While the usability problems didn't result in large productivity increases like the ideation insights, every little 5% improvement helps when you get beyond the V2 of the product.

Attenex Patterns Architecture



attenex

Prototyping

A key part of implementing a visual analytics system is establishing an architecture that combines the content analytics (syntax and semantics) with the visual analytics (Semantics and Pragmatics). The above architecture illustrates how these functions were built into Attenex Patterns. For more on syntax, semantics and pragmatics see: http://www.trinity.edu/cbrown/language/distinctions.html

Purpose/Use	Pragmatics	*Most social*
Meaning	Semantics	*Lexical, logical*
Word external	Syntax	*English Grammar*
Word-Internal	Morphology	*Derivation/Inflection*
Phonemes	Phonology	*Vary for each language*
Speech sounds	Phonetics	*Most physical, universal*

http://www-personal.umich.edu/~jlawler/words/370/Semantics.pdf

Ramana Rao who is now a Stanford professor (http://www.ramanarao.com/) and was one of the founders of Inxight on seeing Attenex Patterns shared that we had solved the holy grail of visualization by using the visualization as an analysis tool as well as a communication tool.

FACTOR10x

90% of Corporate Data is Unstructured
How do you see patterns in the unstructured?

Attenex K&L|GATES

Legal Electronic Discovery

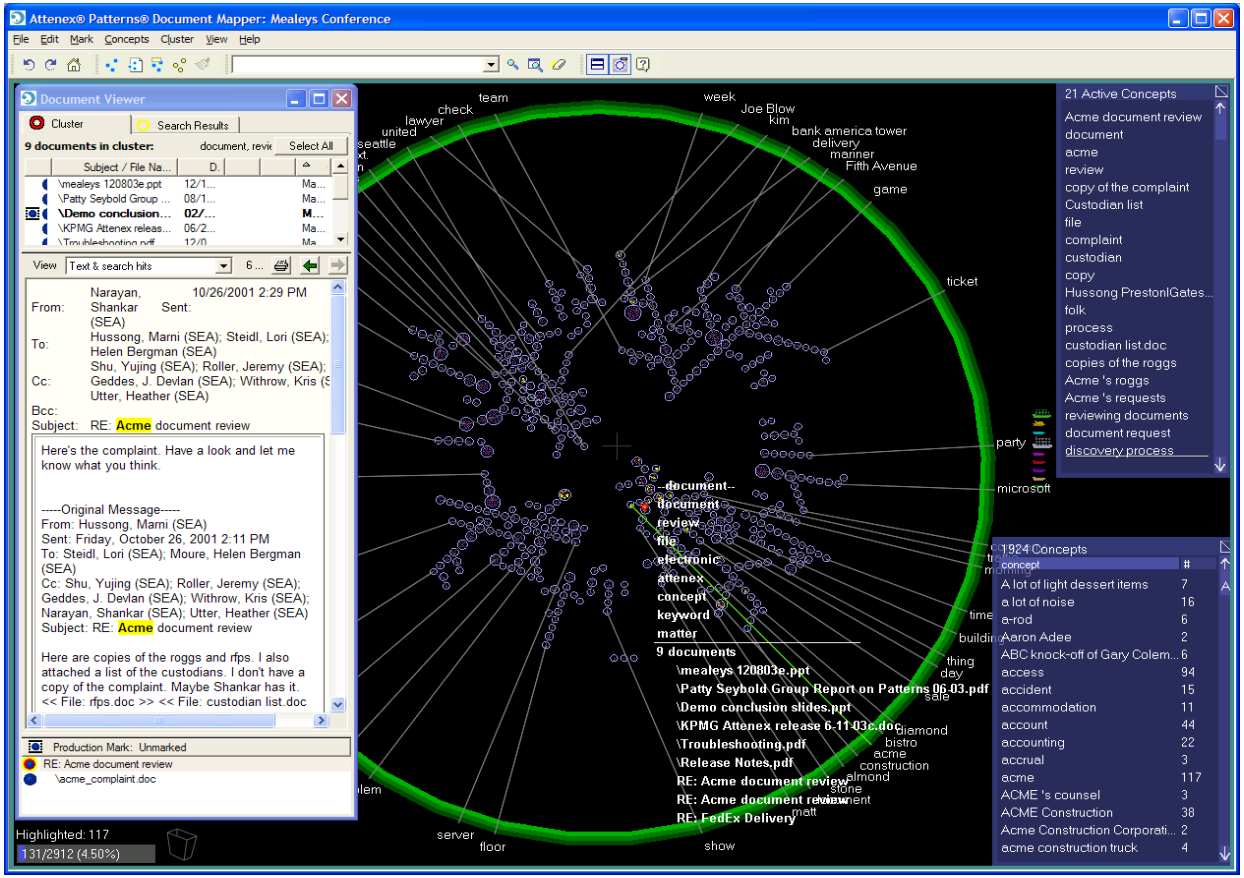
"Attenex Patterns uses visualization as a filtering tool, not the visualization of already filtered information."

Ramana Rao, CTO, Inxight

"Attenex creates structure out of unstructured information in order to JOIN with structured data to improve productivity through multi-aspect visualizations."

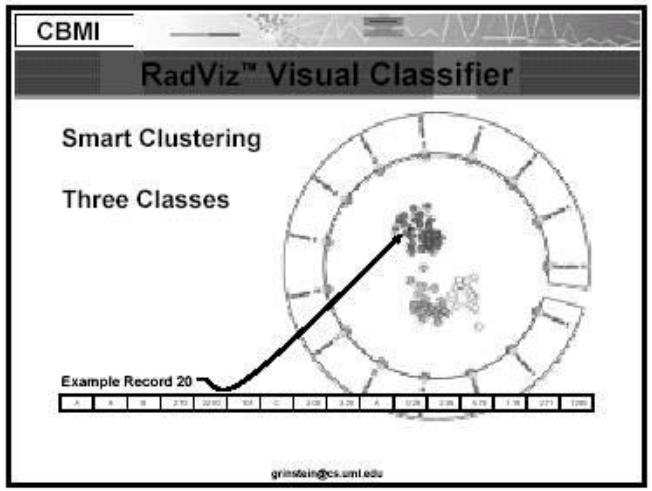
See What Matters - Quickly

5/24/04: Patterns 2.0 Concept Compass



Prototyping

The Attenex Patterns 2.0 release had my favorite visualization feature of all the releases that we did. You never know where good ideas will come from. Three of our developers attended an InfoVis conference in 2003. One of the papers described some research on RadViz a Visual Classifier. The following two images capture the essence of the RadViz classifier.



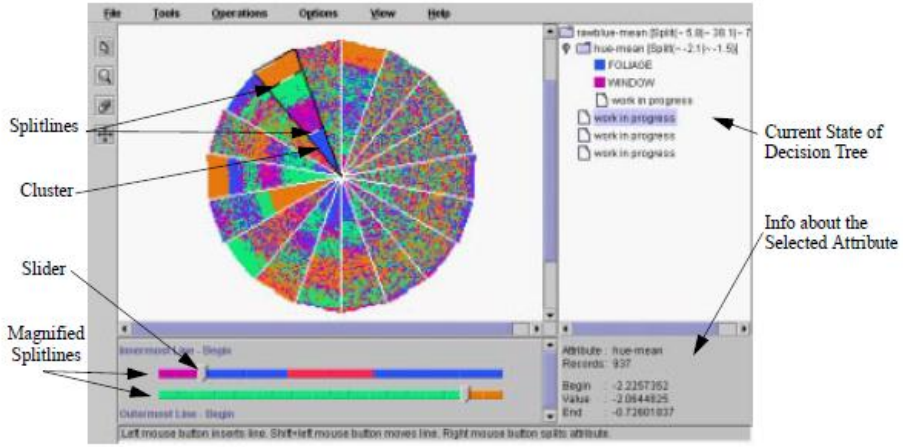


Figure 3. A Screen Shot of the PBC system

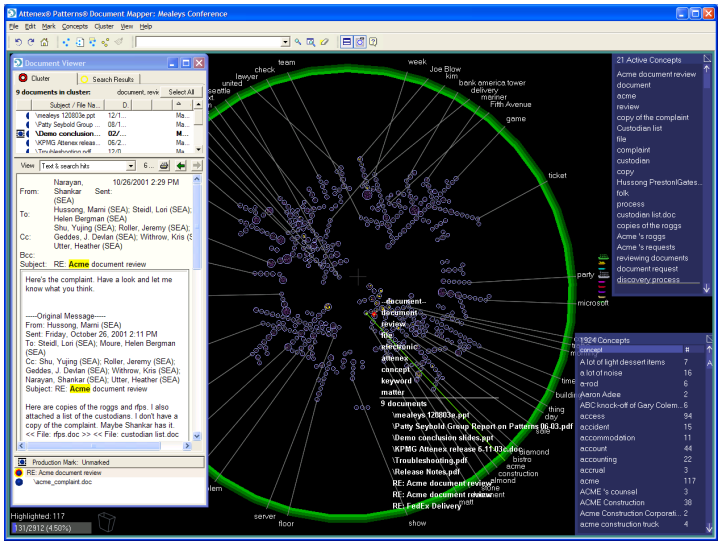
Lynne Evans took these concepts and started developing prototypes of what became the Concept Compass displayed above. After some 20 prototypes and lots of user testing we settled on Concept Compass. To our pleasant surprise we got a full 2X performance increase (that is we went from 10X performance increase over paper document review or page by page electronic review to 12X). While the users couldn't explain why their performance increased, we surmised that the increase was due to always being able to see the Whole as well as the parts. In addition, users reported that as soon as they opened up a new document collection they could determine whether this collection was highly responsive or highly unresponsive based on the kinds of concepts that clustered on the edge of the concept compass.

By having the compass, the layout of the spines was also more understandable. The user could at a glance see what caused the major spine clusters to come together. A hidden benefit of the concept compass is that we were able to display >20% more documents on a screen without confusing the users. It ended up being a way to compact the map.

In addition to making the product more productive, we applied for and were granted two patents out from the work.

Name the Dots

- Snails
- Star Gazing
- Death Star
- Butterflies
- Petri Dish
- DNA
- Staph Infection
- Blood Shot Eye
- Snowflakes
- Airport Terminal
- London Underground
- Erector Set
- Legos
- Alien Circuitry

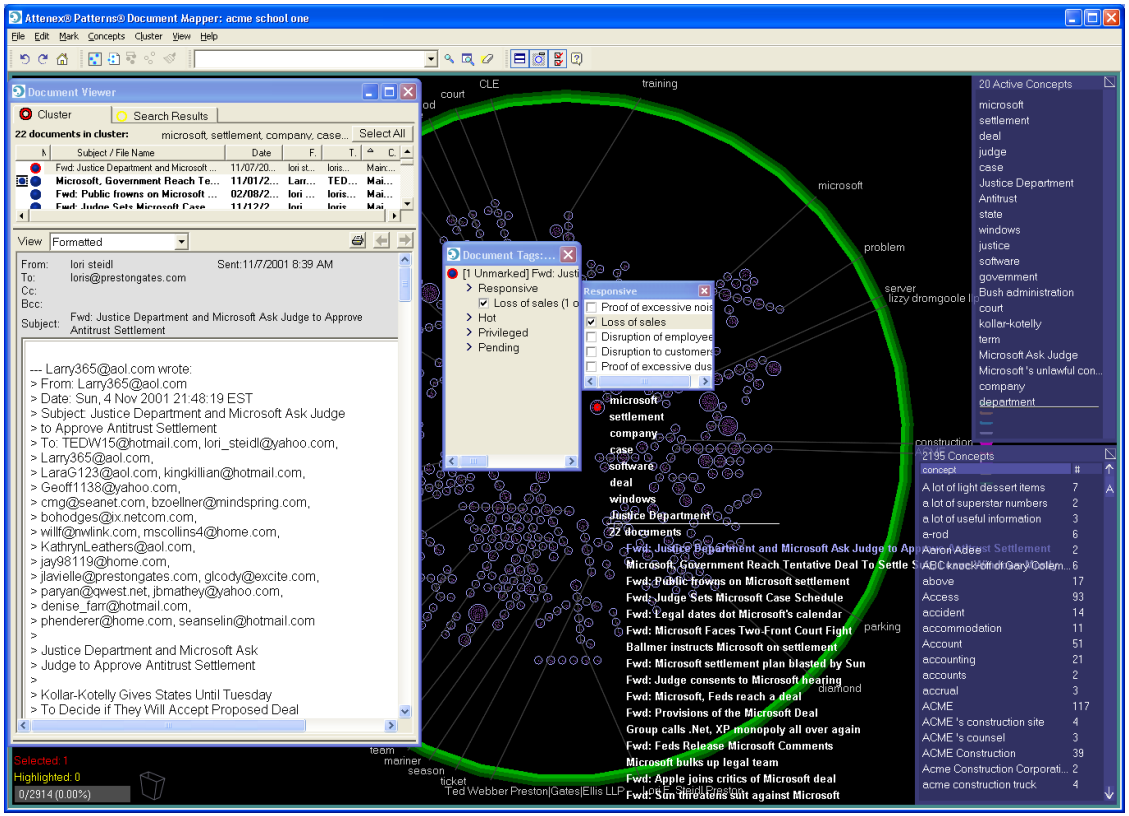


User Research

Another form of user research occurs when you are doing market tests or active marketing of your product. One of the side benefits of having a highly visual product that was very different from the competitors is that the product attracted a lot of attention. One of the challenges of presenting a product at a trade show or a customer demo is that it is awkward or time consuming to profile who is standing in front of you. With Patterns as such a visual product, most potential prospects couldn't help blurt out that the visualization reminded them of something like a "petri dish." With that simple blurting out we had a better idea of their educational background and what kind of metaphors might appeal to them.

As we gathered more customers, we periodically ran contests for our customers to name or apply their favorite metaphor to the visualization. Maybe it is my always being jet lagged but my favorite was always "blood shot eye".

10/8/04: Patterns 2.5 Document Tagging



Prototyping

In late 2004, we released Attenex Patterns 2.5. With this release we added tagging and annotations. We had resisted adding any form of explicit tagging from the product. The more we and our customers used the product, we came to understand that the ability to dynamically manipulate documents and concepts was far better than having a fixed taxonomy or manually tagged information. Most of us in the Western world have come to rely on taxonomies. It is part of our analytic tradition.

However, we finally had to give in and implement tagging. The major reason was that we started using the product for a lot of Hart Scott Rodino Second Requests in anti-trust cases with the DOJ and FTC. As part of these requests the federal government expresses many concerns about the anti-competitive nature of a merger or acquisition. These concerns are numbered and they expect you to not only review the documents but also tag the documents with which concern number the document is responsive to.

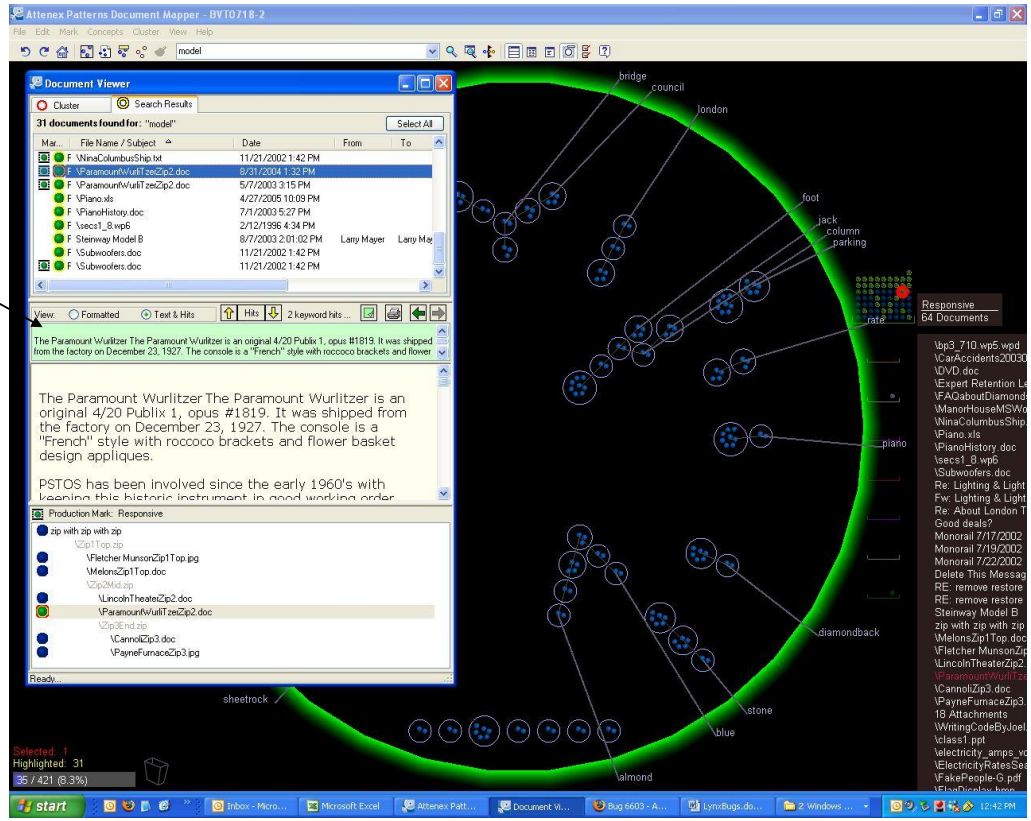
While we felt a better way would be to have the DOJ and FTC be Attenex Patterns users, that process would take years to accomplish. So we finally implemented tagging and annotations.

It took us at least 20 different prototypes to get a version that didn't completely obliterate productivity.

While it is pretty straightforward to add tags, there are several parts to the problem. The schema has to be updated and you have to spend as much time figuring out how to both apply the tag and to search/filter on a given set of tags. In other words we had a whole new set of features that needed to be integrated into our scheme for highlighting and then reclustering.

7/25/05 – Patterns 3.0 Annotations

Annotation



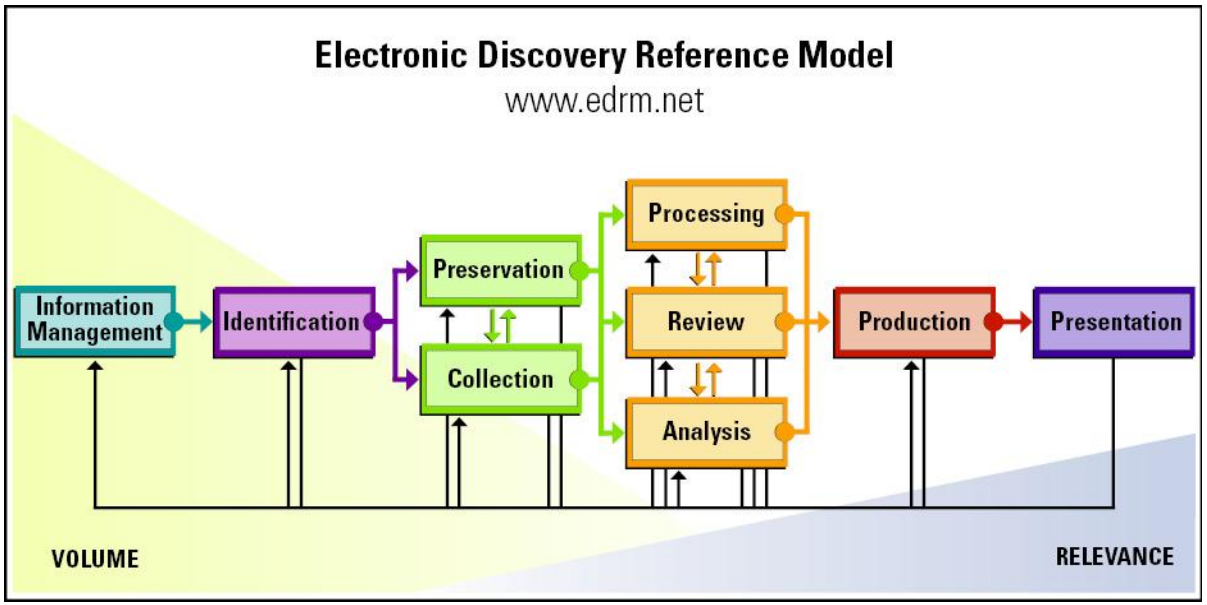
Prototyping

As soon as we added tagging, the users started clamoring for annotations. The 3.0 release featured annotations. The challenge with annotations is how you want to go about searching for the text in the annotation. We now had to have a search all function as well as the option to search only on the actual documents or only on the annotations. We had to be very careful that the annotations were clearly stored outside of the actual documents so that the lawyers could not pollute their defendants actual data.

In the display above you can also see a couple of additional panes added to the Docviewer pane. The top pane has additional metadata on the documents that are in the cluster. The bottom pane has how the documents will actually be produced so that you can see what markings are carried through any email threads (in case an embedded email has a marking of responsive but the rest are non-responsive).

You can also see at the top of the docviewer pane as to whether you wanted to look at the meta data associated with documents in a cluster or documents that were a set created by some search activity.

Establish Industry Standards



28

Value

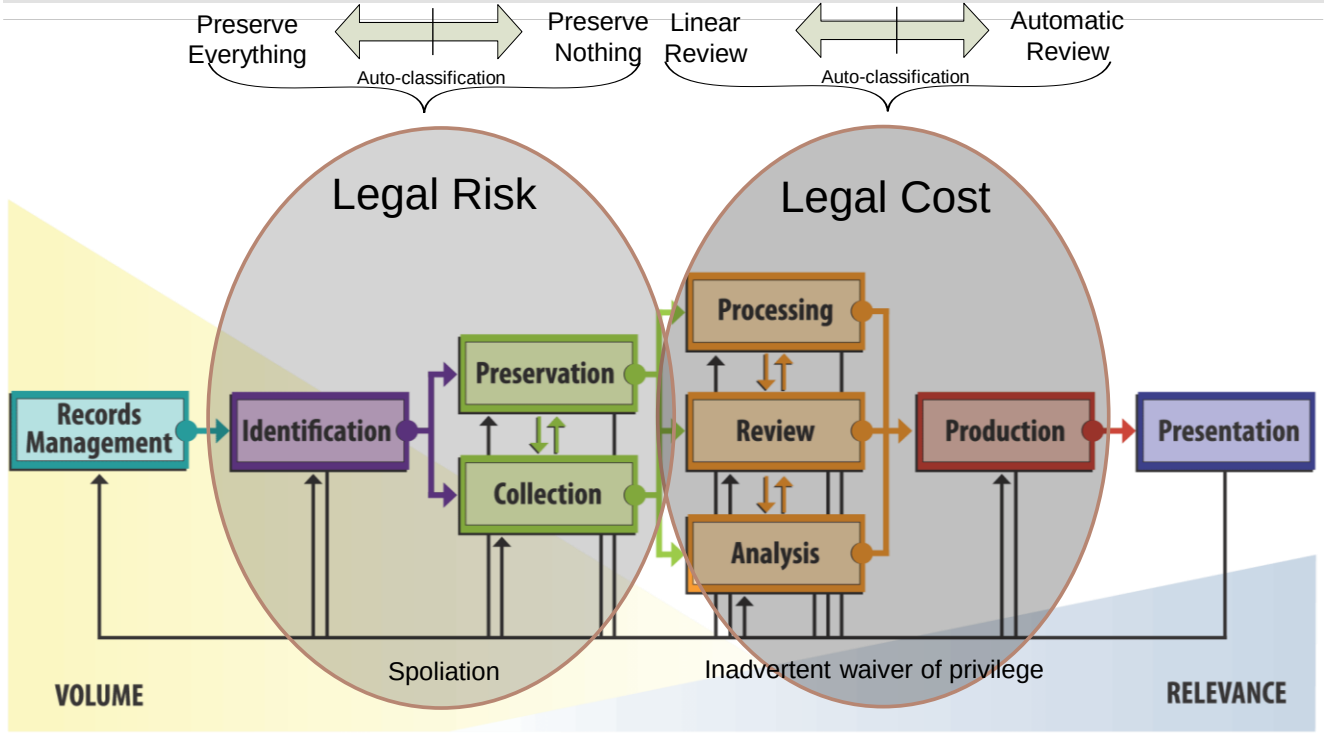
Early on in the process of trying to sell Attenex Patterns, we realized that the industry had a tower of babel when it came to describing what each service provider or technology supplier called different aspects of the eDiscovery process. Since there were over 500 different suppliers at the time, we realized that we needed to establish some industry standards. Attenex was one of the founding members of the Electronic Discovery Reference Model (http://www.edrm.net/) facilitated by George Socha who published the Socha-Gelbman Electronic Discovery Survey (http://www.sochaconsulting.com/).

Through this process we developed a model and standard terminology so that when we presented the capabilities of our product we could be assured that our customers new what we were talking about.

This aspect of human centered design is often overlooked – the ability to communicate the innovations that you develop so that your customers can understand them. In a new market segment, you often have to create an "independent" standards body to facilitate the language structures and models to talk about what it is you are doing. Without this language, only the very early adopters will buy.

Having a common language also helps make the selling function be more efficient as you are not having to teach the customer the important concepts.

The Opportunity



Value

While having the model was a good start, it was also important to communicate where in the process the risks and costs showed up. As we mentioned at the start of the user research we looked for the bottlenecks in the overall review process and focused on where the large chunks of money were being spent. This diagram nicely captures the steps where the costs are incurred and also where we focused. We gladly let others focus on where the risk was (early part of the process) because relatively speaking there was so little money spent in that arena.

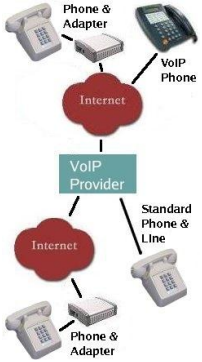
From a value standpoint any time you can find the part of the process where there are high costs and low risk (or liability) that is a great place to focus.

eDiscovery Fodder - Corporate

2006 – 20 Petabytes

BELLSOUTH

500 Terabytes of email



2002 – 2 Petabytes

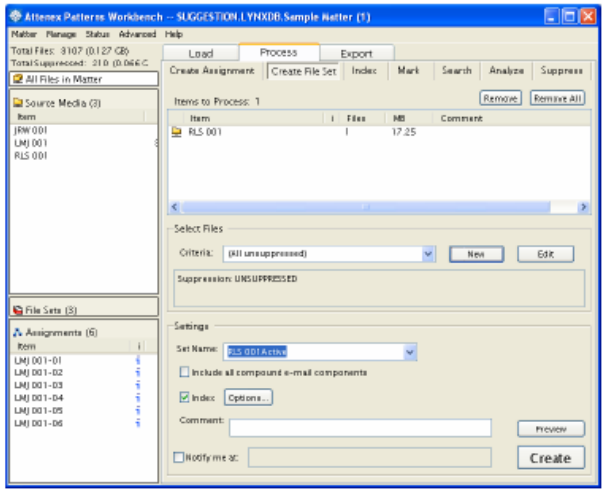
Value

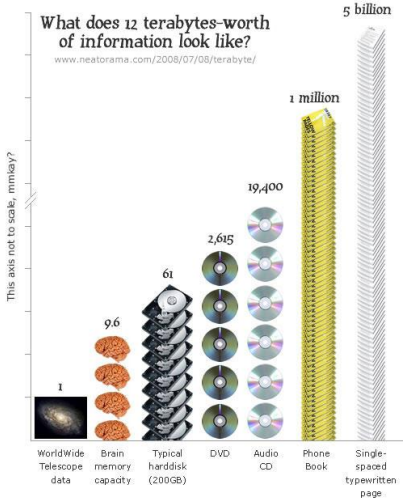
For most of us it is hard to comprehend how large the amount of electronic information is that is subject to requests for production in litigation. The operating telephone companies are amongst those with the largest amount of online always accessible digital information (although in these times of big data I am sure that Google has far more online information).

To give you some idea of how much digital information there is, if you had to print out a Terabyte of documents and place them in bankers boxes, the boxes would occupy four Sears Towers (Chicago).

The School of Information Sciences at UC Berkely regularly studies how much digital information there is:
http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/

7/25/05 Patterns 3.0 Workbench





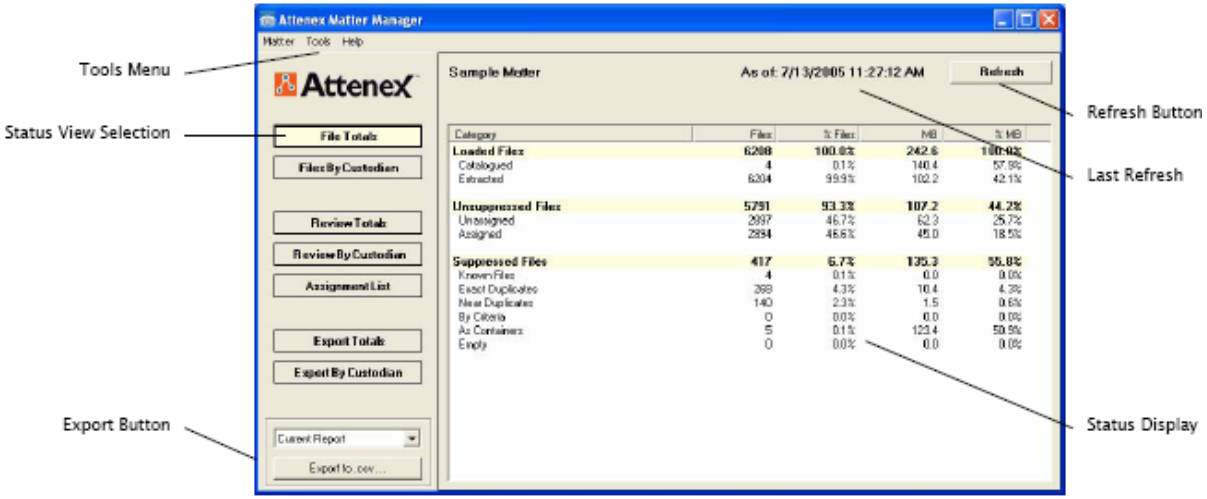
Prototyping

While the Attenex Patterns product is used by 1000s of reviewers every day, the user interface to the backend product is only used by the "data wranglers" who acquire the documents and process them. This tool has much less time devoted to the user interface since the users will be well trained IT Specialists, but has an extraordinary amount of time devoted to improving the performance. So the prototyping of algorithms is key to improving the backend. We went through as many prototypes for the backend system as we did the front end; they just aren't as visible.

7/25/05 Patterns 3.0 Matter Manager

Matter Manager Main Window

Once you open a matter, the Matter Manager main window is displayed.



User Research

However as we did more work with our users we realized that there was a persona that wasn't being well served – the project manager for a matter. So we went back and did the user observation and video ethnography on the matter managers and produced this module at V3.0. This tool functions as the dashboard for a given matter.

See What Matters:
Attenex Patterns Visualizations

WHAT?

Semantic Network:
Concept map displays documents based on content similarity

WHEN?

Event Network:
Timeline visualization maps documents based on time

WHO?

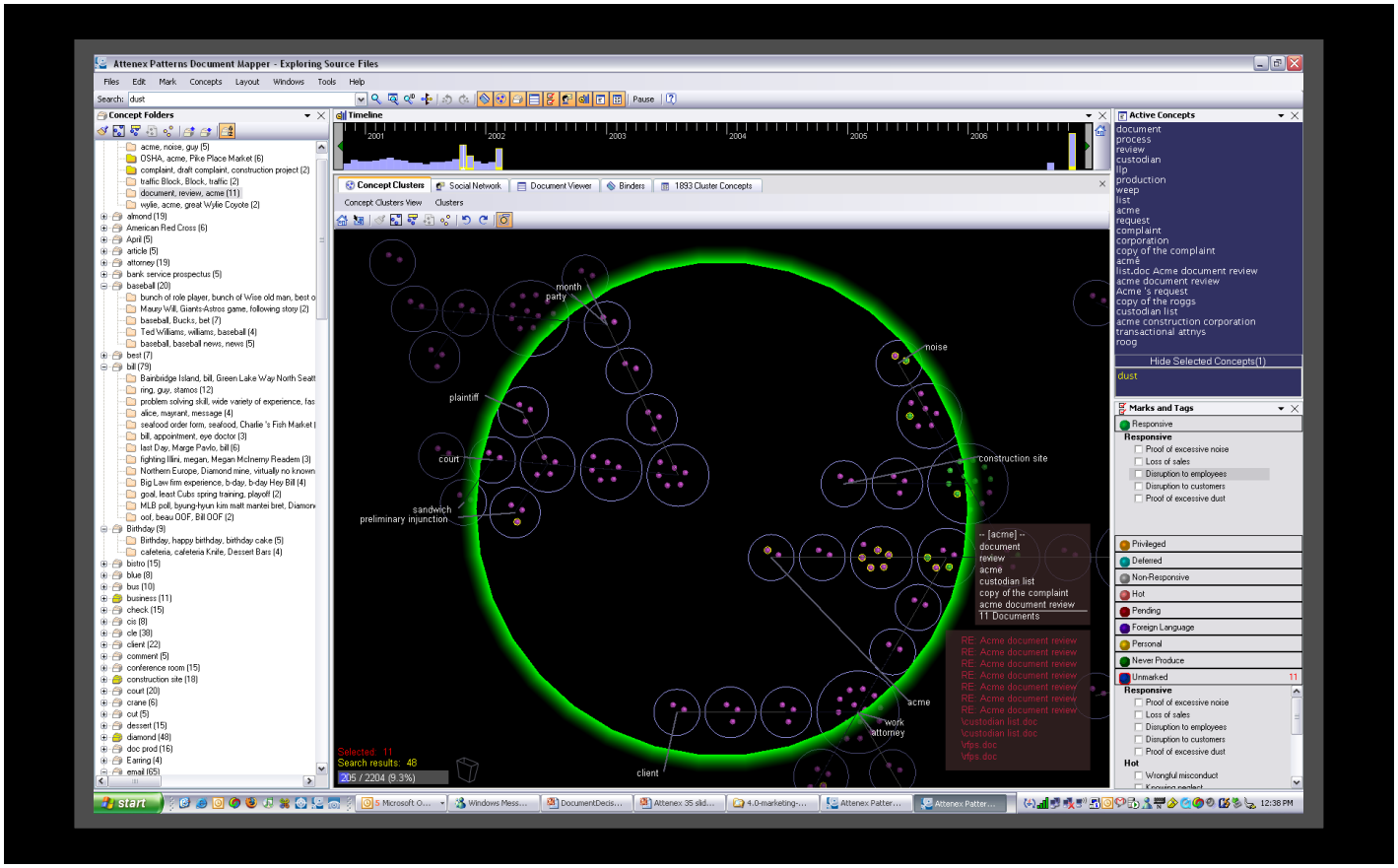
Social Network:
Social network displays relationships in email communications

User Research

As we expanded the types of matters that we processed, our users asked for additional visualizations. In one anti-trust case it was critical to see who was communicating with whom internally as well as which music entities different divisions were communicating with externally. Looking at the emails themselves it was difficult to tell the social networks because of the sure volume. So we implemented a social network view. Similarly many cases wanted to know when communications occurred so we created the event network diagrams. Here are examples of the numbers of variables we plotted in each view. In addition each of the separate windows had referential integrity, that is, when you performed an action in one network view the other windows would change appropriately.

- Semantic Network
    1. Concepts - this is the core variable plotted in this space for which the other variables are related
    2. Clusters - documents which have similar concepts
    3. Spiral within the cluster - within a cluster, documents are plotted based on their similarity - high to low
    4. Spines - clusters which are related to each other by one or more concepts
    5. Proximity - clusters which are closer together are more related than those which are farther apart
    6. Concept Compass - how spines and spine concept are related to each other (spine concept by number of clusters which have that concept)
    7. Highlighting - which clusters and documents have a selected concept
    8. Recluster on Highlighting - change how clusters are created (with all the ways that one can create highlighting)
        1. This is where the variables can really add up as highlighting can be generated in so many ways including the referential integretity of a document or a relationship being highlighted in the social network or timeline views).
    9. Recluster on Document - change how clusters are created - the whole of a documents concepts, not just the six concepts that cause documents to cluster initially
    10. % of Doneness - shown in two ways - the bottom left count and the set aside trays
    11. Decision code (an color) - Responsive, Non-Responsive etc
    12. Tagging - given that you can search and reorient things based on the tag
    13. Docviewer ordering - you could probably go even crazier if you included the number of ways that you could order selected documents within the top frame of the DocViewer
    14. [NOTE: As a general rule I only talk about the variables 2-6 when I give a talk.]
- Social Network
    1. People View (left hand side of the email @)
    2. Domain View (right hand side of the email @)
    3. Connections between (can think of this as one variable or three)
        1. People
        2. Domains
        3. In Quicksilver, sub-domains - domains are a container
    4. Directionality of view - do messages flow in one direction or both (indicated by color)
    5. Volume of messages between people and domains (Quicksilver)
    6. Quantity of Dollars (Fraud within Quicksilver - if the message is a financial transaction, rather than an email)
    7. Degrees of separation - Kevin Bacon indicator which gets at the hops between people\
    8. Highlighting - all the ways that can be highlighted in Social Net and the other views
        1. Select a relationship - and who else was involved in those communications
        2. Select and organize by the number of sent or received messages
        3. Select and organize by one or more people (or domains)
    9. [NOTE: As a general rule I only talk about the first four when giving a Patterns demo and then add the next three when I gave Quicksilver demos.]
- Event Network View
    1. Date message or document was sent
    2. Number of messages per unit time
    3. Highlighting

Grouping Documents into Clusters and Concept Folders



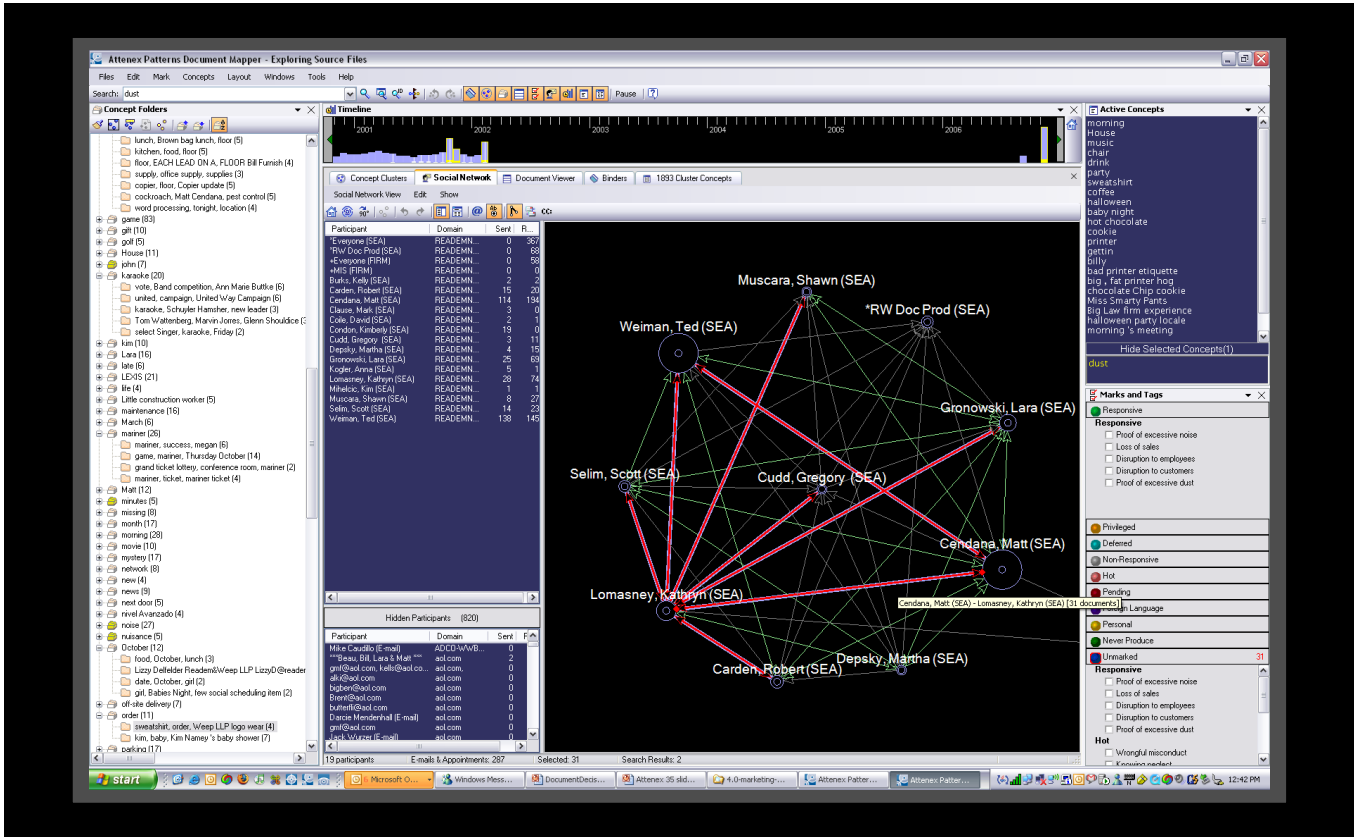
Prototyping

With great reluctance we finally bowed down to user demand and added a “foldering” view pane (the common form of Windows and Microsoft Office views of file systems). This pane is for the left brain analytics who just can’t wrap their heads around a visual interface. We did quite a bit of user testing and found that indeed productivity did not improve (in fact it decreased slightly) with the addition of this view.

The screen shot shows the number of different panes that a user could have open with each of the buttons in the right hand panes serving as filtering tools to highlight documents in the main display that meet the criteria.

You can also see that we’ve moved to iconic menu buttons in the second row of the display rather than text based buttons. These buttons turned on or off different view panes as well as doing the reclustering on highlighted etc.

Examining Social Networks with Document Collections



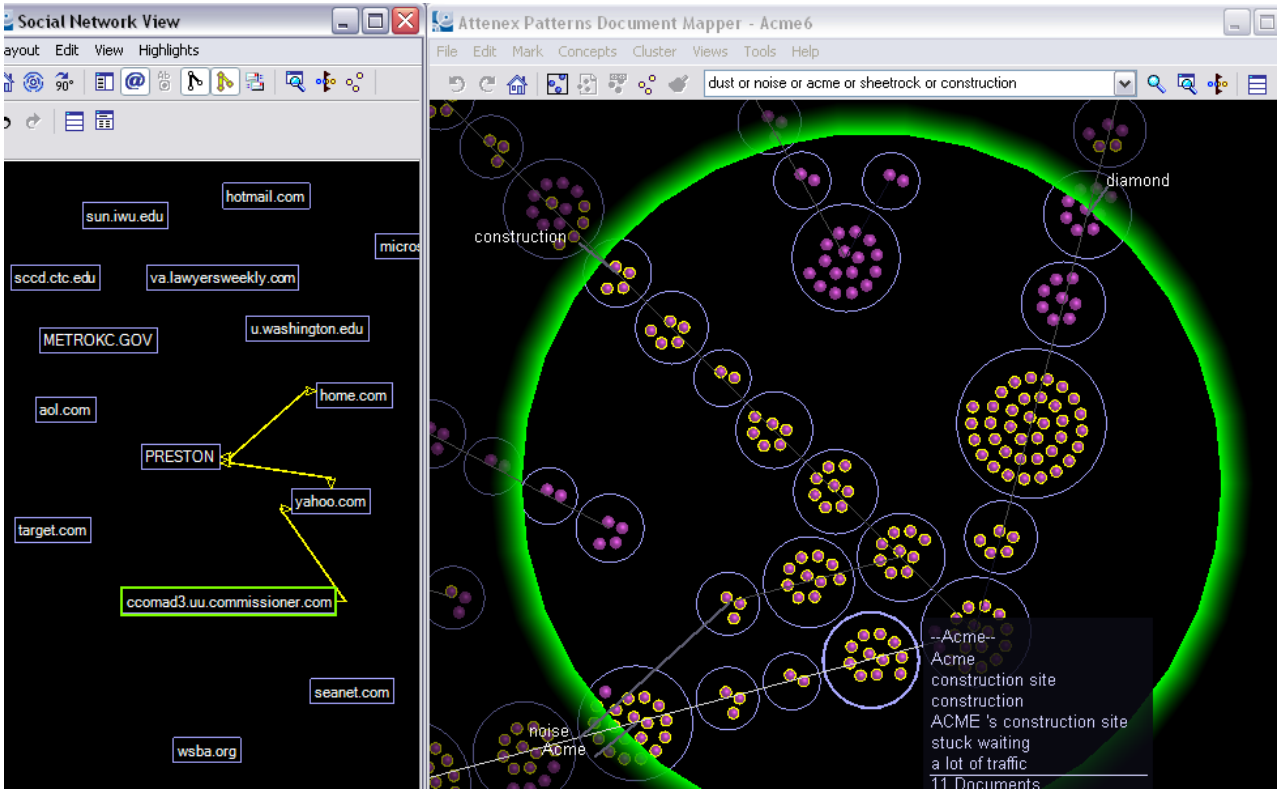
Prototyping

Where the previous screen shot showed the main pane as the semantic network, this screen shot shows an example of a social network. In the middle you can see two panes where you have the selected participants emails that you want to look at. The middle pane then shows the links between individuals (the left hand side of an email address – left of the @ sign). Each line has arrowheads which illustrate which direction the communications flowed. Looking at the arrowheads you can quickly tell who is sending out email messages to a list. The width of the line indicated the relative amount of communication between the individuals.

The size of an individuals node illustrates the relative volume of emails in total that individual had.

When you click on a link (arc) between the nodes a list is generated of all communications between those two individuals (show in a semantic network pane). If you click on the node itself you will see all of the emails between that individual and the other selected indivisuals.

Attenex® Patterns® - Combining Views for Rapid Understanding



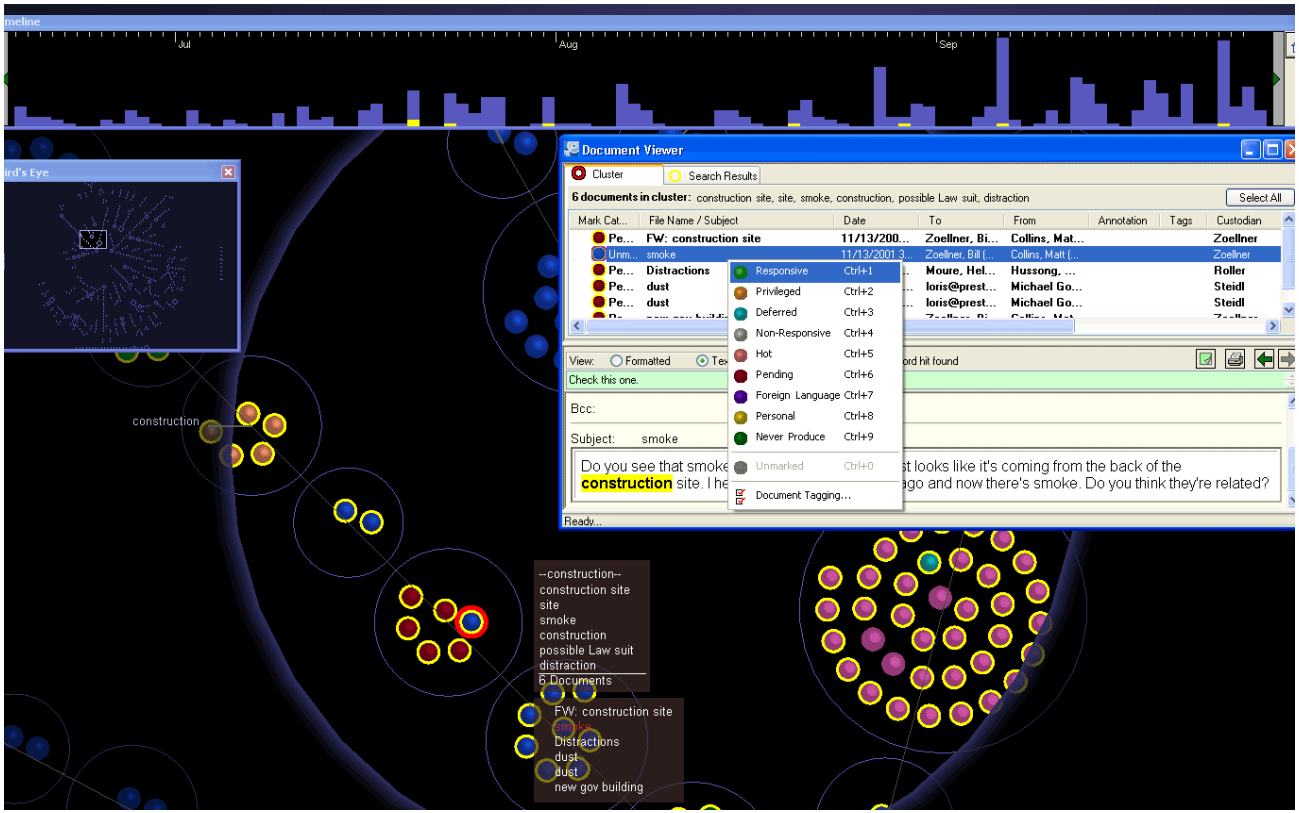
Prototyping

This screen shot illustrates another form of social network – looking at the communication between organizational entities (the text to the right of an @ side in an email address). By having both the social network graph and the semantic network graph (right hand pane), you can look at what concepts and documents are passing between the two entities.

If you click on an arc on the left hand pane, all of the documents that are between those entities will be highlighted with a yellow ring around the dot in the right hand pane.

If you click on a document or a document cluster in the right hand pane (semantic network) the corresponding entities that generated that message will be highlighted in the left hand pane.

2/1/07 V4 Semantic Network and Event Network



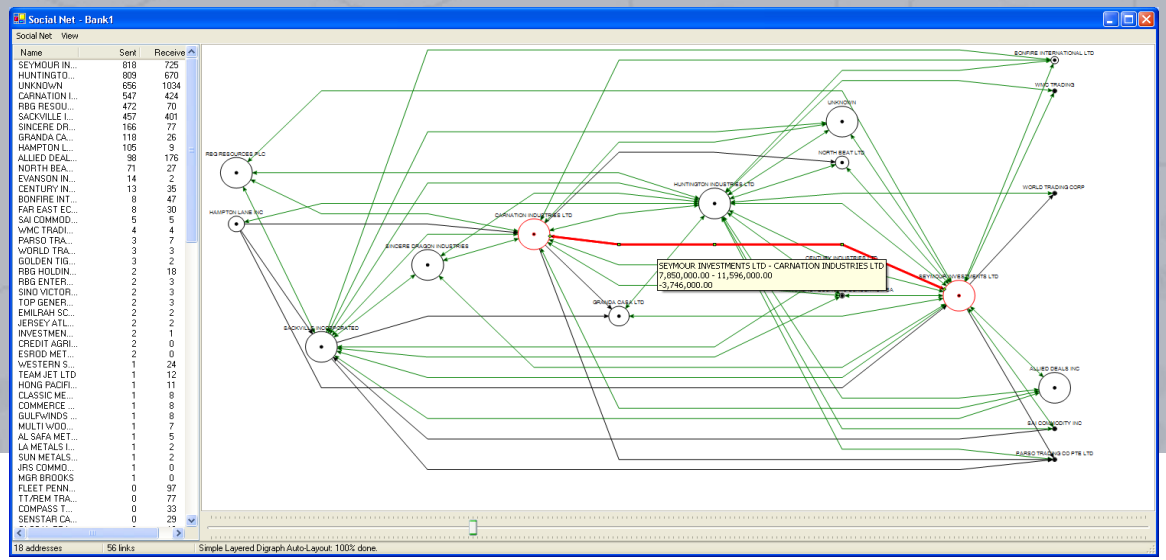
Prototyping

Another combination of network maps is to combine the semantic network view and the event network view.

The event network is the top pane and shows the volume of emails by date. As you select documents in the semantic network view when those documents were created or sent are displayed in the event network view. Likewise if you were to click on a bar (or set of bars) in the top pane, the documents/messages that occurred on those dates would be displayed in the semantic network view.

Visualizing Financial Transactions
Suspected Wire Transfer Fraud (Same tool set)

- Visualize suspected wire transfer fraud
  - One way or two way transactions
  - Symmetric or asymmetric dollar transactions
  - Number of transactions per customer or supplier
  - Immediate return of wire transfer



Prototyping

KPMG, one of the Attenex partners, asked us if we could build a prototype that would help them visualize fraud which occupied a lot of the time of their forensic accountants.

We found that we just needed to add a few capabilities to the standard social network view to visualize the flow of dollars. As it turns out most financial transactions consist of a date, time, to, from, subject line and message fields (like amount) much like the structure of an email. So all we needed to do was add a summing function for the amount field to add to the capabilities of a financial transaction network.

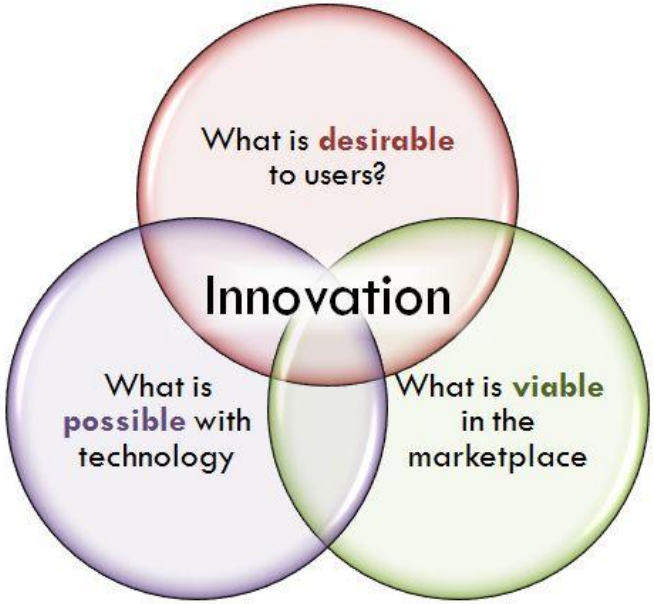
In the example above we looked at >50,000 wire transfers from a client of Citibank that they suspected of doing fraudulent transactions to inflate their revenue numbers. The slide above shows you the rules for how you identify potentially fraudulent transactions. The yellow box looks at the relative flow of dollars in both directions. Some $7M flowed in one direction and some $11M flowed in the other.

The large bubbles in the diagram illustrate the sham companies that the fraudulent company set up to move money around.

With this prototype we were able to see in 10 minutes what took 2 forensic accountants 120 hours (@$500 per hour) to solve.

Unfortunately, KPMG was unwilling to change their billing model to accommodate the extreme productivity that the visual analytics provided.

Essence of Commercialization



http://www.ideo.com/work/human-centered-design-toolkit/

34

Value

IDEO has done an excellent job of capturing the essence of human centered design in their HCD Toolkit: http://www.ideo.com/work/human-centered-design-toolkit/ . An adjunct to the toolkit is the excellent book by the CEO of IDEO, Tim Brown, *Change by Design.*

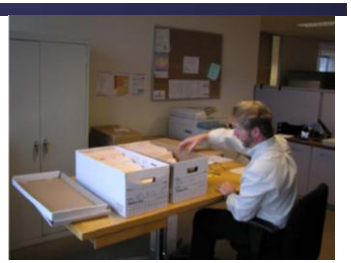
Alan Cooper in his book *The Inmates are Running the Asylum* provides these definitions of the three aspects of HCD – capability (or possibility or feasibility), viability, and desirability.

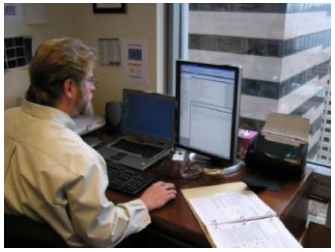
> Larry Keeley of the Doblin Group has created an intriguing conceptual model of three primary qualities in high-technology business. Keeley calls the first quality *capability*, and it is what technologists bring to the party. They ask, "What are we capable of? What is possible?" Engineers must know what can and can't be built. A product can't be a success unless it can be built and made to work.
>
> Keeley calls the second quality *viability*, and it is the contribution of businesspeople. They ask, "What is viable? What can we sell?" Business executives must know what can and can't be created and sold at a profit. A product can't be a success unless it can support a growing company.
>
> Because all successful high-technology businesses need a balance of both of these qualities, the tension between their constituents is strong. Businesspeople are totally dependent on technologists for their ability to create things that work. And technologists are totally dependent on businesspeople to provide them with the tools for their efforts. This makes for an uneasy symbiosis.
>
> Keeley calls the third quality *desirability*, and it is what designers supply. They must ask, "What is desired? What do people want?" Designers determine product behavior that will make people happy and satisfied. A product can't be a long-term success unless it delivers power and pleasure to the people who must actually use it.
>
> Design takes a product that can be built and performs well, and that can be distributed and sold profitably, and makes it a success by making it into something that *people really want*. This third leg brings stability and converts an interesting technological achievement into a long-term success.

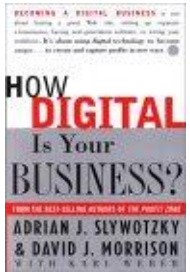
The first part of this briefing book focused on the capability and desirability aspects of Attenex Patterns. The remainder of this briefing book will focus on viability.

Relentless Pursuit of Productivity

- Using Slywotzsky's imperatives, observe knowledge workers
  - Move from guessing what customers want to knowing their needs;
  - Move from getting information in lag time to getting it in real time;
  - Move from burdening talent with low-value work to gaining high talent leverage.
- Identify productivity outcome metric
  - Attenex: Increase document decisions per hour
- Develop prototype(s)
- Test for productivity outcome





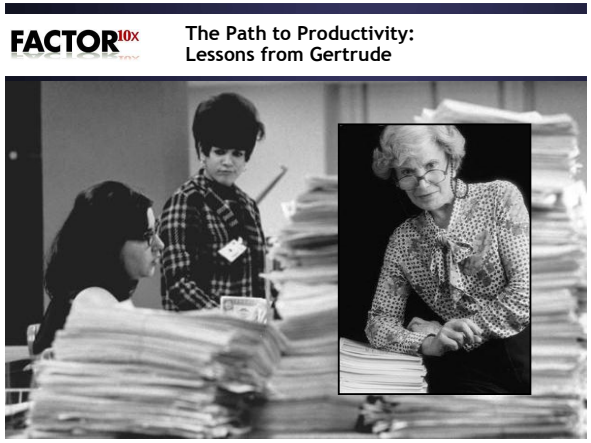


Value

An important part of providing value to an enterprise purchaser is to solve for at least one of Adrian Slywotzky's knowledge worker imperatives (see *How Digital is your Business?* http://www.amazon.com/Digital-Your-Business-Adrian-Slywotzky/dp/0609607707/ref=sr_1_1?s=books&ie=UTF8&qid=1318525848&sr=1-1).

With Attenex Patterns we focused primarily on the third imperative – "move from burdening talent with low-value work to gaining high talent leverage." The lawyer reviewers are all highly educated professionals (college plus law degree). Yet, the work of review is mind numblingly boring.

By moving to the concept map, we allowed the reviewers to think strategically about a collection of documents. By organizing the documents much like a paralegal would do, a reviewer could be assured that all of the documents in a given cluster were closely related (as opposed to randomly related by the order they hit an email system or a file server).



"For some context, I thought I'd share a little "history of document review."

"In 1973, the law firm of White, O'Reilly and Walter found itself buried in a document review of nearly one million pages. It was at this time that a thought-leading paralegal named Gertrude Buckets revolutionized document review by instructing her team to help the attorneys by grouping documents into piles of "like" information.

"The attorneys quickly realized that they could make more quickly and accurately determine what documents were relevant by having documents organized for them prior to review. " From an Attenex Corporate slide presentation.

Growth Partner

How can you grow your business?

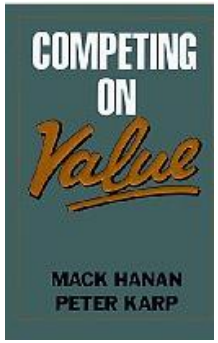
You cannot.

You can only grow someone else's business. His business growth will be the source of your growth. By growing, he will force growth back upon you because he will want you to grow him again.

The businesses you can grow have a name. They are called your major customers. Their growth must be the objective of your business. The capabilities you require to grow them must be your asset base.

Growth requires a partner. A growth partner is a special kind of customer. He is a customer whose costs you can significantly reduce or whose profitable sales volume you can significantly increase. In one or both of these ways, you can improve his profits. This is the basis for his growth. It is also the basis for his contribution to your own growth. As the two of you grow each other, you will become mutually indispensable.

If you cannot grow a customer, you cannot partner him. You can continue to do business with him, buying and selling, but the maximized profits of growth will elude both of you. If all your customers are buyers instead of growers, you will be a slow-growth or no-growth business. None of your customers will be growing you because you will not be growing them.



38

Value

Mack Hanan's advice is critical for any startup company. For a long list of legal reasons, Attenex could not sell directly to other law firms or corporations. We needed a layer of service providers (or growth partners) between us and the end customer. These service providers became our growth partners. One of those growth partners (FTI Consulting) eventually acquired Attenex for $91M.

Productivity with Visual Analytics

- Impact of Attenex Patterns
  - 2001: 300GB took 200 attorneys 1 year (Anti-trust Matter – Pre Attenex)
    - *Total Matter Cost to client: > $18M (0% technology, 100% labor)*
  - 2003: 300GB took 100 attorneys three weeks (Energy Fraud Matter)
    - *Total Matter Cost to client: > $6M (30% technology, 70% labor)*
  - 2006: 300GB took 65 attorneys 2.5 days to review (Board Investigation)
    - *Total Matter Cost to client: ~ $2M (50% technology, 50% labor)*
- 2007 Revenue for Attenex and **Partners**
  - Attenex: $30 M
  - Partners: > $300 M

Value

While we were able to pretty quickly arrive at a visual analytics prototype that would increase productivity by 10X, it took us much longer to find a business model that would work. In the legal industry the predominant business model is the billable hour. So if we didn't change the business model, we would be pitching to our customers "if you use our product we will reduce your revenues by 90%." While the answer was obvious in the end, some form of fixed pricing instead of billable hours, it still took us a while to come up with a good model. Fortunately, Preston Gates had 10 years of billing information that we could use to model a reasonable pricing approach. However, we wanted to make sure that we did value based pricing so we could resist as long as possible commodity pricing. The following discussion comes from our original business plan.

Under this cost structure, a firm will be able to pass on a 30 percent cost reduction to its clients while maintaining or even increasing its total revenue.

Currently, initial review services cost $30-40 per megabyte of information processed. Firms maintain a margin of roughly 50 percent on these services. Thus, the cost to the firm is $15-20/Mb. STM produces a 10X productivity gain, meaning the cost to firms to process information is reduced to $1.50-2.00/Mb. By being conservative and assuming that there is only a 5X productivity gain, the firm could charge $4/Mb. Thus, even with STM processing fees of $4/Mb, the firm will maintain or slightly decrease profits even as it reduces client fees by 20 to 25% percent, as shown in the tables below.

Effect of STM on Costs and Fees, based on existing $30/Mb Discovery Costs			
	Fee to Client	Cost to Firm	Profit to Firm
Current System	$30	$15	$15
With STM	$24	$8.00 (incl. $4/Mb STM processing fee)	$16
Change	- 20%		+7%

Effect of STM on Costs and Fees, based on existing $40/Mb Discovery Costs			
	Fee to Client	Cost to Firm	Profit to Firm
Current System	$40	$20	$20
With STM	$30	$8.00 (incl. $4/Mb STM processing fee)	$22
Change	- 25%		10%

While these tables show the pure cost reduction of applying STM to the discovery process, they do not account for associated administrative savings to the firm related to accomplishing the same amount of work with fewer workers in less time.

Viability Progression

Costs

Revenue

Profits (EBITDA)

Valuation

Exit – what multiple of revenue can you achieve

Time to Revenue

2

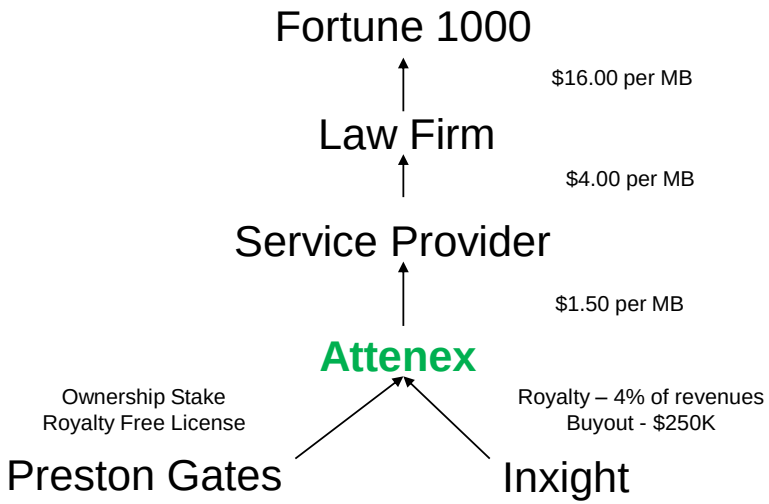
Value

Over the years, it has taken me a long time to realize what the key goal is for an entrepreneur – increase one's valuation.

While you are faced with trying to understand valuation the minute you try to raise money to fund your venture (what is your pre-money valuation?), nobody ever spends time going through what leads to a good valuation and what can you do every day to increase your valuation. There are lots of resources to help you with how to manage costs, revenues, and profits, but very little on how to increase valuation.

An excellent book on what leads to valuation is *Early Exits* by Basil Peters. If we had this book before we sold Attenex to FTI Consulting we could have received an additional $30M in valuation.

Attenex Example
Customer's Customer Value Chain



How does each member of the value chain make profits?

What is the value proposition for the whole and each component of the value chain?

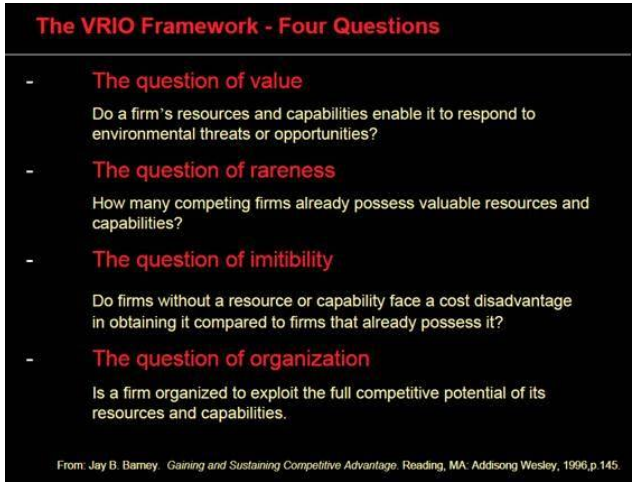
How will profits move over time in this value chain?

Value

From a business strategy standpoint, one of the things we had to do at Attenex is understand where we fit in the value chain from suppliers to the end customer for our product. The above diagram represents the value chain that we were in at Attenex. While Michael Porter developed the original value chain concept, he didn't spend any time on the questions posed above. Due primarily to legal reasons, we had to place ourselves in the position displayed above.

The numbers to the right represent the relative pricing for each of the layers in the value chain. All of these prices are value based pricing, that is they represent prices based on the value to the customer rather than on the cost to produce the product (commodity pricing).

Another framework for thinking about where to place yourself in the value chain comes from Jay Barney's VRIO framework (see *What I Didn't Learn in Business School: How Strategy works in the Real world.* The following images capture a definition of VRIO and then an example in Barney's book.

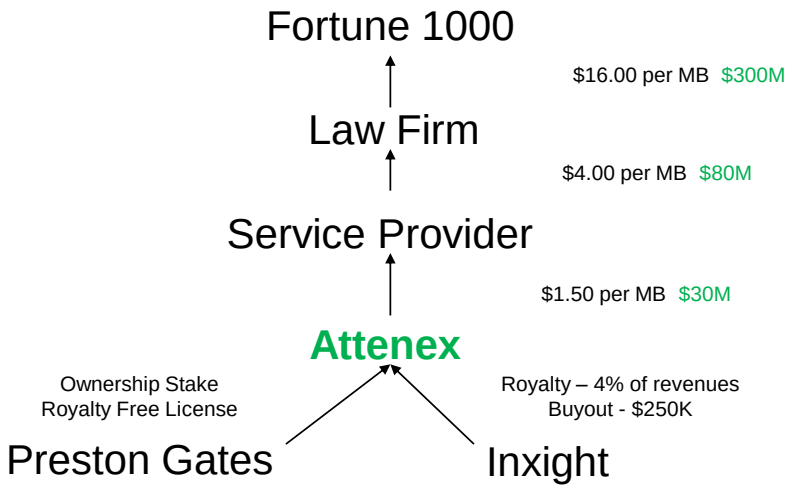


Justin's VRIO analysis of the Plastiwear shirt value chain

Stages in the value chain	Are there opportunities in this stage for HGS to increase revs/decrease costs?	Does HGS have any any unusual skills in executing this strategy?	Will it take other firms significant time to imitate this strategy?	Advantage potential?
R&D				
–License Plastiwear	?	Experience with Plastiwear	Maybe—patent + experience	?
–Alternative uses	?	Experience with Plastiwear	Maybe—patent + experience	?
Acquire raw materials				
–Volume purchasing	Probably	Probably not	?	At best, parity
Fiber manufacturing	Probably, yes (if shirt demand exists)	Probably not	—	At best, parity
Fabric weaving	?	?	?	?
Cut and assemble	Probably, yes	Probably not	—	At best, parity
Deliver shirts	Probably, yes	Probably not	—	At best, parity
Retail sales				
–Open retail stores	?	?	?	?
–Internet sales	?	?	?	?

A firm's core competencies are those activities it engages in that (1) create economic value, (2) are rare among competitors, and (3) competitors find difficult or costly to imitate.

Attenex Example
Customer's Customer Value Chain



How does each member of the value chain make profits?

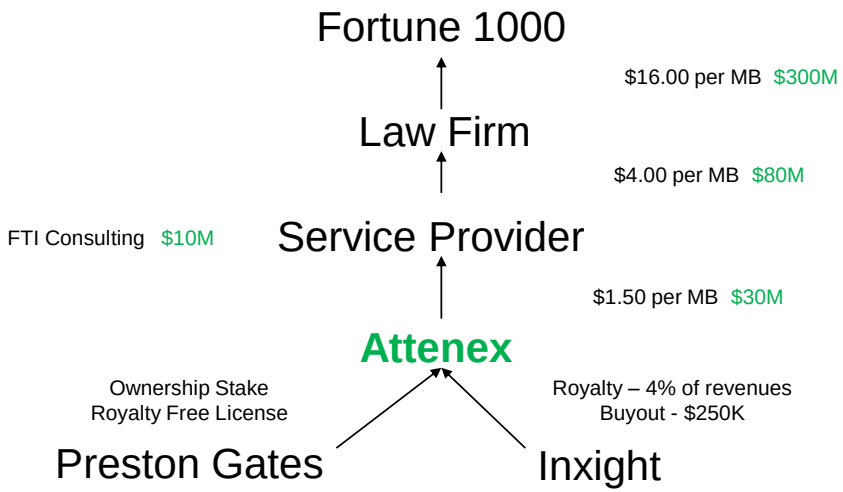
What is the value proposition for the whole and each component of the value chain?

How will profits move over time in this value chain?

Value

The green dollar figures in the above diagram represent how much revenue each of the layers in the value chain derived from Attenex Patterns. While we grew to $30M in revenue our service providers were receiving $80M a year while the law firms were receiving $300M a year. At each layer, the gross margin on these revenues was greater than 50%.

Attenex Corporate Strategy
Sell to FTI, Expand into Value Chain, or Enter New Market (Patents) - 2007



How does each member of the value chain make profits?

What is the value proposition for the whole and each component of the value chain?

How will profits move over time in this value chain?

Value

Looking at the above numbers, what would be your next move at Attenex. Should Attenex sell itself (for $91M) or enter a new value chain (like the patent analytics space) or move to a different place in the value chain (like the review services business). Attenex chose to sell itself. Then two years later FTI chose to go into the review services business (Acuity service offering http://www.ftitechnology.com/Products-Services/Software-and-Services/Acuity/Acuity.aspx). The Acuity business is regularly referred to in the FTI SEC 10K and 10Q filings and their annual report as their fastest growing and most profitable business.

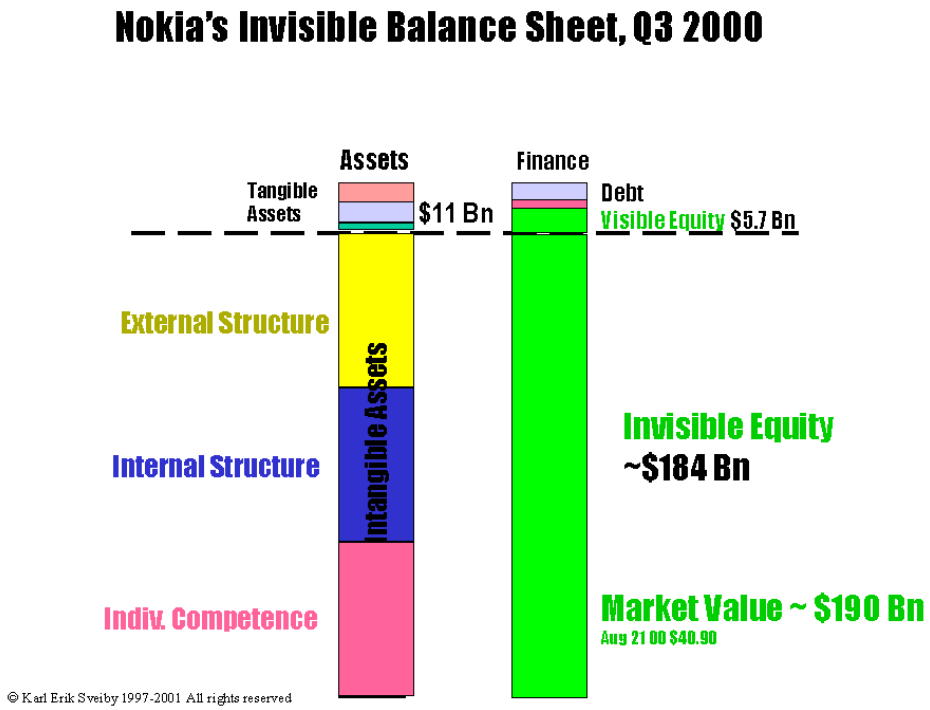
Intellectual Capital & Product Offerings

- Intellectual Capital
  - Talent (Human Capital) – the talent base of the employees
  - Structural Capital – non-human storehouses of information
  - Relationship Capital – the knowledge embedded in business social networks – both customers and suppliers along with strategic networks
- Product and Service Offerings
  - Contingent Services – eDiscovery Review Services, Tier 1 experts, brokering legal technology
  - Professional Services – management consultants, eDiscovery Strategy consultants, Law Firm Partners
  - Content – IP and contract knowledge base, patent and trademark database, lawyer database, case law and statute database (FastCase)
  - Software – Content Analyst, Attenex Patterns, Summation, HotDocs, Wikis, Blogs, Enterprise Search Tools (Stellent, Oracle, Google), Custom development, Website Development

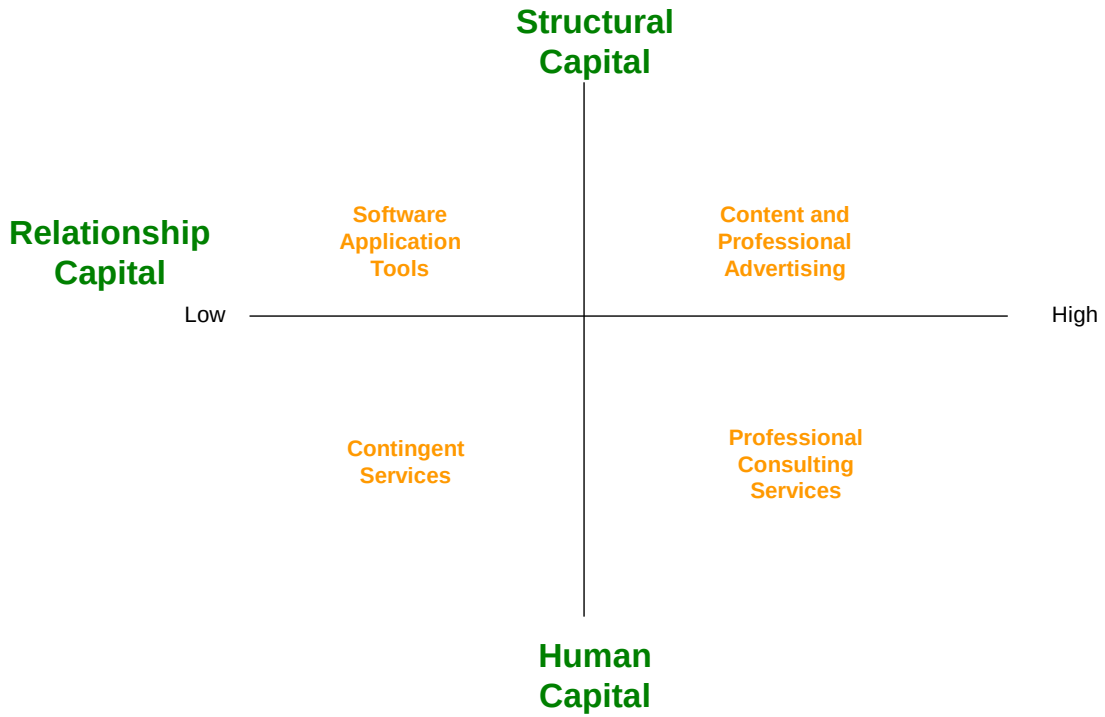
Value

Another part of understanding valuation is understanding why for public knowledge based companies the stock valuation is so much higher than traditional finances indicate with their book value. Tom Steward, Leif Edvinson, and Karl Svieby have written extensively on Intellectual Capital over the years as a way of explaining this difference.

From Svieby - http://www.sveiby.com/articles/InvisibleBalance.html



Intellectual Capital



Value

To help executives with the progression up the curve to focus on Intellectual Capital and Valuation I put the above 2x2 quadrant together on the relationship between intellectual capital and product offerings.

The vertical orientation of the matrix is a representation of how much the product or service requires a human being to deliver the product or service.

The horizontal axis of the matrix is the degree of relationship capital that is involved in the product or service. To the right of the matrix requires a high degree of relationship capital, while the left hand side is low on the relationship side.

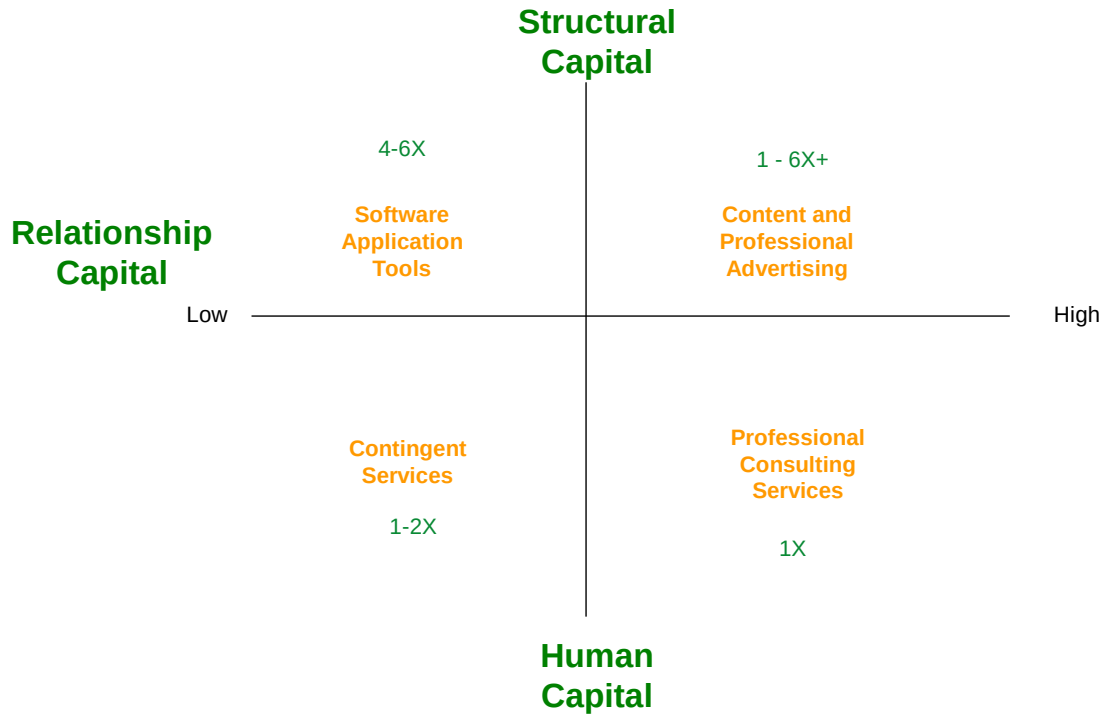
The lower half of the matrix are services that require human beings to deliver. Contingent services are provided by relatively low skilled, fungible labor that is most often out sourced. The labor is paid for typically by the hour and the humans only work when there is work to do (they are not on salary). It is not necessary that a particular person be assigned to the task, but rather just somebody with the needed basic skills. In the case of Knovel there are content preparation tasks that can be performed by contingent labor.

Professional services are typically provided by full time employees who are knowledge workers. The accounting firms, law firms and consulting firms like Anderson and McKinsey and large scale services organizations like IBM have most of their workers as professional services employees. These employees typically take quite a while to train in what is needed for their particular product offering and it is difficult to keep these employees fully occupied due to the vagaries of the project cycle. These employees are either billed at high hourly rates or at project rates.

The upper half of the matrix is technologically based and can be delivered with no human beings involved. Software applications and tools are high on structural capital, but low in relationship capital. That is in order for software to be a product it is relatively generic and does not have a relationship to the user. Alan Cooper put this nicely with his comment that "Software Forgets".

On the upper right is structural capital that has a high relationship value. Perhaps the best example is Google and its targeted Adwords and Adsense where the ability to charge higher dollars for the ads is directly related to how much Google knows about the user who is accessing the content. The more the content is targeted to a specific user the more the content producer can charge for the content whether in Google (through advertising) or in a specialized expertise areas like the Gartner Group (http://www.gartner.com/technology/research/) or the eDiscovery Journal (https://ediscoveryjournal.com/research/).

Intellectual Capital



Value

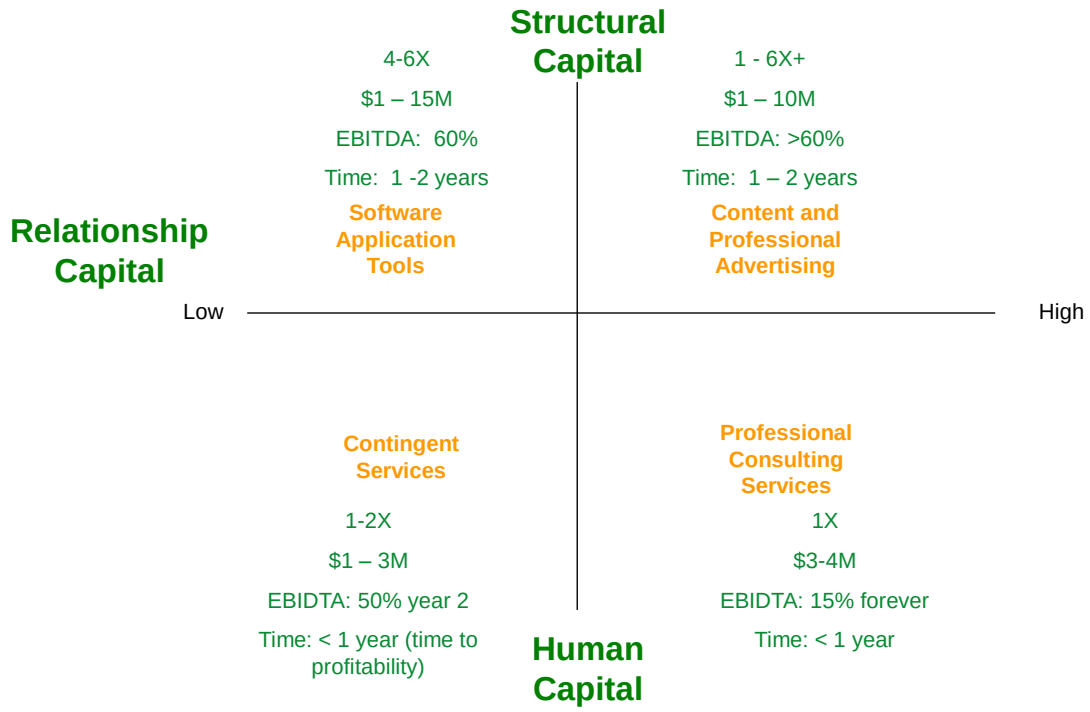
As we look at what is implied by each of the quadrants, we see that there are different multiples of revenues associated with each of the quadrants. The software applications on average provide the highest multiples (4-6x) of revenue because they are almost pure structural capital and require very little in the way of human capital to customize for a particular user. While it may take a lot of investment to create the first copy of a software product, the 2nd to nth copy cost virtually nothing with electronic distribution (see p. 8 on Design, Build, Distribute, Intervene, Operate in "Is designing software different from designing other things?"). Since there is little in the way of relationship capital with software, not a lot of customizing and "hand holding" is required once the software is released.

If we look at the two human capital quadrants we see that the valuation is roughly equal to the revenue. The services require human beings to deliver and those human beings leave every evening. While each employee in a professional services firm can generate more revenue per hour than a person in a contingent labor business, the profits are relatively fixed at 20% or less. In a contingent labor business, because the person is only working when there is actual revenue to be had often times the profits can be quite substantial which is why a contingent labor business has the possibility of a higher revenue multiple for its valuation.

The most interesting quadrant in recent years is the upper right with content. Until the advent of Google, the content business was generally a multiple of 1x of revenue. This is the area of TV, Radio, magazines, and print media. Usually a content company was limited to a 1-2x multiple because each company had to produce its own content (the editorial side) along with selling and generating advertising. Typically these companies had a lot invested in the production and distribution side of the business (TV stations, TV broadcast towers, four color high volume printing presses, physical distribution of magazines). Google stood everything on its ear by being completely digital and not requiring an editorial side as they just indexed and used other people's content. In the age of Google, content can command a 6+ revenue multiple for valuation if the company is all digital and has figured out how to monetize the content in multiple ways (going beyond just advertising – see multi-sided business models).

Content when combined with software (particularly when using the r2NDA concept) can command even higher valuation multiples of revenue.

Intellectual Capital



Value

This slide looks at some of the cost sides of the equation and what level of profit might be generated (EBITDA - http://www.investopedia.com/terms/e/ebitda.asp). It also gives an indication of the time it takes before meaningful revenue can be generated and the level of investment required to get that revenue.

4-6X
$1 – 15M
EBITDA: 60%
Time: 1 -2 years

If we take a look at the numbers for software what this information says is:

- Multiple of revenue that can be generated for valuation: 4-6X
- Amount of money invested to get to a product: $1-15 Million
- EBITDA: 60%
- Time it takes from start of investing to realizing revenue: 1-2 years.

A more complete definition of EBITDA:

What Does *Earnings Before Interest, Taxes, Depreciation and Amortization - EBITDA* Mean?
An indicator of a company's financial performance which is calculated in the following EBITDA calculation:

$$EBITDA = Revenue - Expenses\ (excluding\ tax, interest, depreciation\ and\ amortization)$$

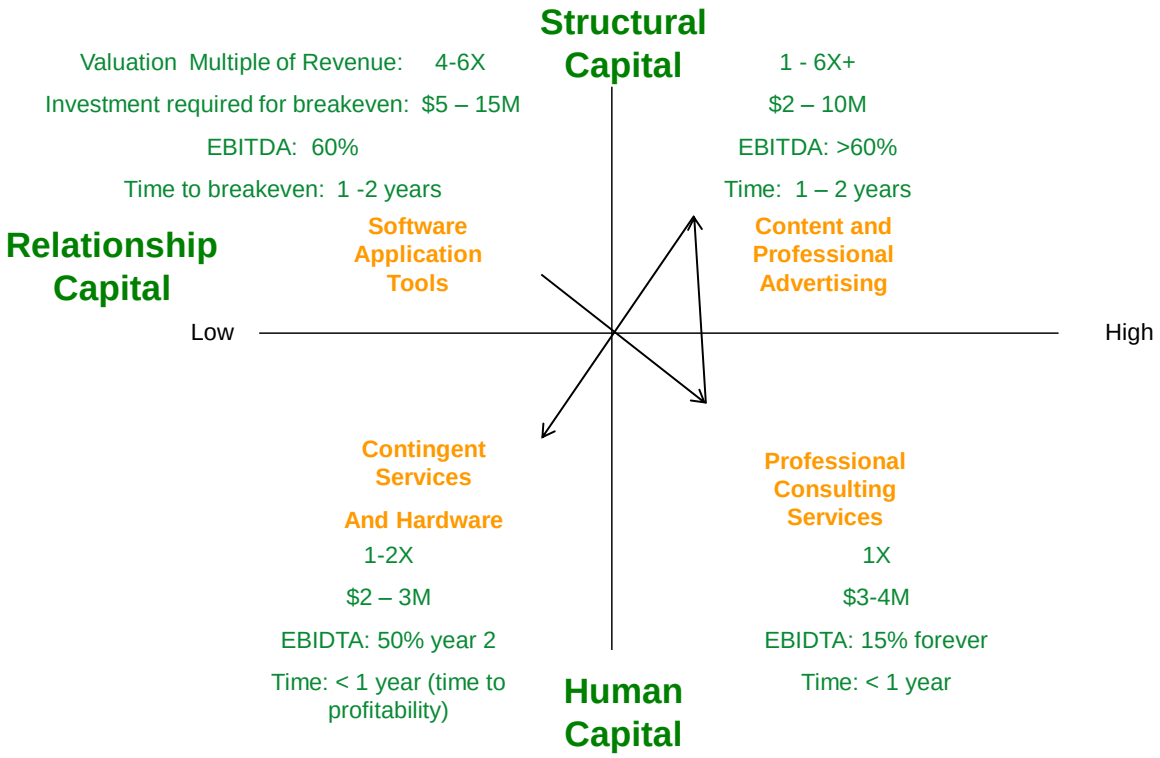
EBITDA is essentially net income with interest, taxes, depreciation, and amortization added back to it, and can be used to analyze and compare profitability between companies and industries because it eliminates the effects of financing and accounting decisions.

Investopedia explains *Earnings Before Interest, Taxes, Depreciation and Amortization - EBITDA*
This is a non-GAAP measure that allows a greater amount of discretion as to what is (and is not) included in the calculation. This also means that companies often change the items included in their EBITDA calculation from one reporting period to the next.

EBITDA first came into common use with leveraged buyouts in the 1980s, when it was used to indicate the ability of a company to service debt. As time passed, it became popular in industries with expensive assets that had to be written down over long periods of time. EBITDA is now commonly quoted by many companies, especially in the tech sector - even when it isn't warranted.

A common misconception is that EBITDA represents cash earnings. EBITDA is a good metric to evaluate profitability, but not cash flow. EBITDA also leaves out the cash required to fund working capital and the replacement of old equipment, which can be significant. Consequently, EBITDA is often used as an accounting gimmick to dress up a company's earnings. When using this metric, it's key that investors also focus on other performance measures to make sure the company is not trying to hide something with EBITDA.

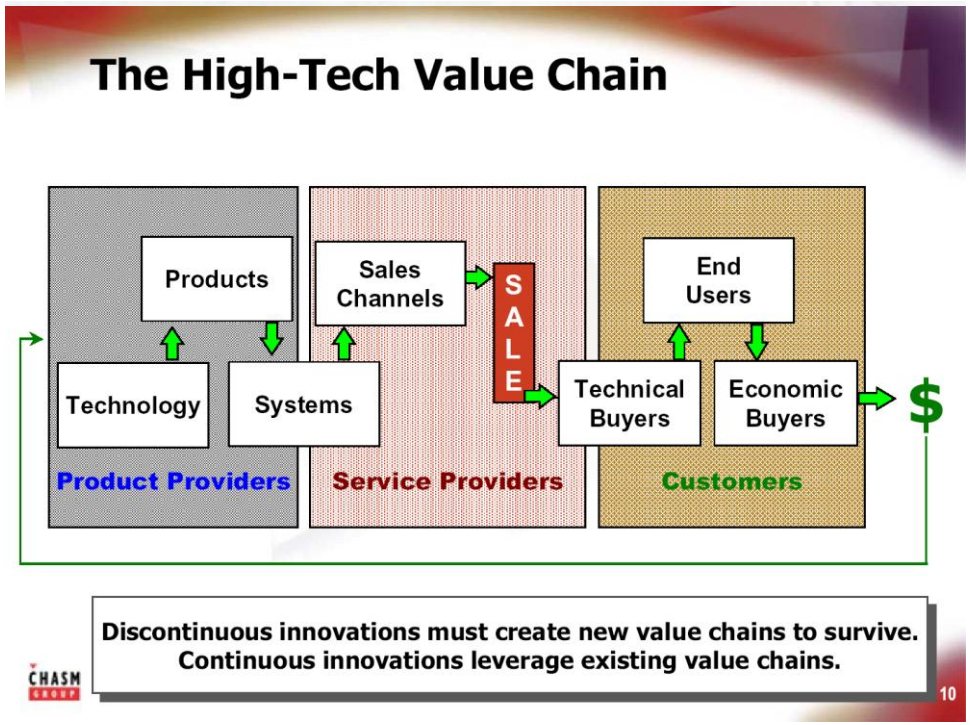
Intellectual Capital



Value

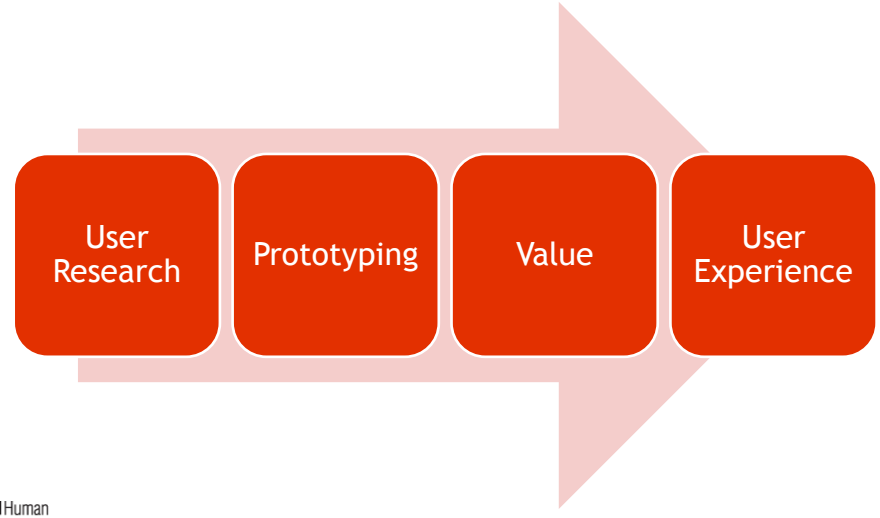
While most business strategists recommend that entrepreneurs have relentless focus, the digital business needs to think about operating in multiple quadrants in the above model. The above progression illustrates that you might start in the software application and tools quadrant and then add some professional services around your product as a way to get paid to figure out how you should enhance your software and then you move to the content quadrant to get a higher valuation and then you might move to contingent services (or even hardware like Apple does with iOS iphone and iPad and Google recently did by buying Motorola) to generate even more content.

Geoff Moore presents this in his book *Gorilla Game* with the evolution of a startup:



FACTOR 10x

HCDE 518
Course Overview



HCDE Human Centered Design & Engineering

35

Questions?

Skip.Walter@factor10x.com
206-200-6777

FACTOR 10x

Extreme Productivity by Design