Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Overview of Stages of PLR Preparation and Associated Tasks

Anthony Trippe
Managing Director – Patinformatics, LLC
WIPO Regional Workshop on Patent Analytics
Intellectual Property of the Philippines (IPOPHL)
Manila, Philippines – 5 December 2013

Patinformatics, LLC®

Data Driven Decisions

Patent Strategy and Analytics Services

STAGES OF PLR PREPARATION

PLR Preparation

- Planning – covered in previous talk
- Performing the search
- Preparing the data to be analyzed
- Statistical analyses to include – to be covered in subsequent talk
- Additional statistical analyses to consider including – to be covered in subsequent talk
- Additional analysis types to consider including – to be covered in subsequent talk
- Writing and publishing the report and accompanying data – covered in previous talk

Performing the Search

- Using the Linear Law of Patent Analysis as a backdrop for thinking about searching, there are two elements that need to be considered, the first involves generating a query that will provide the right data for addressing the business question behind the PLR, and the second involves producing the proper output and formats for ensuring that the data can be analyzed in the next step of this process.

Performing the Search

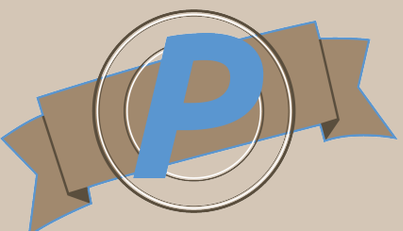
- Determining the type of searching to be done
- Are you looking for pharmaceutically relevant compounds?
- Are you looking for designs or at mechanical devices?
- Are you interested in plants or traditional medicines?
- Are you looking at genetic sequences or biomolecules?

Performing the Search

- Determining which databases to use
- Some databases contain information that can subsequently be used to identify data in other databases that are configured to allow for the exporting of information that will be used in the analysis steps
- The following fields should be available for export: Patent Assignees, Inventors, Application Date, Priority Date, International Patent Classifications, Forward Citations, Backward Citations and Patent Status

Performing the Search

- Sharing the strategy
- The search methodology section of the PLR contains a written description of the various methods used. The breadth and depth of the approach can be clearly seen from the outlined steps, in a fashion that can be understood by the average reader of the document, but without overwhelming them with jargon and scripting
- The details, important to a patent information professional who might be tasked with updating or replicating the work, can be found in the appendix.



Preparing the Data to be Analyzed

- Field cleanup and grouping – covered in Tasks section
- Family or invention reduction
- Is manual review required for precision?
- Determining which year field will be used
- Generating technology categories
- Reconciling forward citations – discussed during practical exercises section

Family or Invention Reduction

- Multi-country filing bias, can occur if a reduction of the set to be analyzed is not performed
- It has been suggested that a One Document per Invention (ODPI) approach , where all inventions from a primary country, the United States for instance, are retained, provides a middle ground for eliminating multi-country biases while ensuring that investment is properly represented
- Many analysts use INPADOC Families

Is Manual Review Required

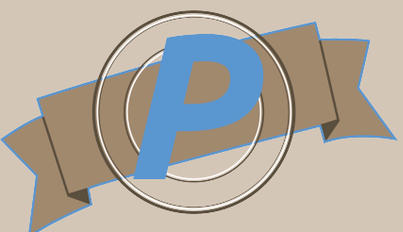
- Greater than 90% recall, and 70% precision are suggested before accurate analyses can be generated
- Even with advanced searching tools it is sometimes impossible to produce a query that will provide at least 70% precision
- There are occasions where 70% precision is not high enough and precision levels of closer to 90% are required for trustworthy results
- Especially true with smaller data collections, were errors will have a larger impact than they would with sets numbering in the tens of thousands.

Determining which Year Field will be Used

- Application year provides a closer approximation to when the research was performed
- But creates a dip in most recent years based on 18-month publication cycle or time it takes to grant
- Publication year does not generate a dip since patents and applications are always publishing
- Don't have to explain sudden downward trends to the clients

Generating Technology Categories

- Technology categories are sometimes identified using the patent data itself, for instance, with classification codes, but ideally they should be generated based on input from a subject-matter expert based on an industry standard view on how approaches are categorized
- Using a market or industry-based approach to creating categories will make it easier for the clients of the PLR to identify with the technology and apply it to the environment that they are already comfortable with.



Patinformatics, LLC®

Data Driven Decisions

Patent Strategy and Analytics Services

TASKS ASSOCIATED WITH PLR PREPARATION & PATENT ANALYTICS

Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Detailed Discussion – Landscaping Tasks

Technique

Data Cleanup & Grouping of Concepts

Definition

Manual or automatic standardization of terms within a data field. List cleanup is required in order to produce statistically relevant results

Utility

Grouping allows synonymous terms to be combined together so that their true value in a data set can be accurately assessed

Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Data Cleanup

	A	B
20	BRUSHTIME PRODUCTS, INC.;BOOKER, Winifred, J.	
21	BRUSHTIME PRODUCTS, INC.;BOOKER, Winifred, J.	1
22	C&C. LTD.	
23	C&C. LTD.	1
24	CHEN, Chunmei	
25	CHEN, Chunmei	1
26	CHIN CHUZAN;沈 仲山;CHIN SHUNRYO;沈 俊 良;CHIN CHUN-CHIUNG;沈 俊	
27	CHIN CHUZAN;沈 仲山;CHIN SHUNRYO;沈 俊 良;CHIN CHUN-CHIUNG;沈 俊	1
28	CHOI, Byeong Gap	
29	CHOI, Byeong Gap	2
30	CJLION CORP.;C&C. LTD;LEE, Eul Kyou;CHOI, Il Gyu	
31	CJLION CORP.;C&C. LTD;LEE, Eul Kyou;CHOI, Il Gyu	1
32	Colgate	
33	COLGATE PALMOLIVE CO	2
34	COLGATE-PALMOLIVE COMPANY	3
35	COLGATE-PALMOLIVE COMPANY;FONTANA, Jose Eder;LEMOS, Edilberto;PERNA, Fernando;FOCASSIO, Paulo	1
36	COLGATE-PALMOLIVE COMPANY;PATEL, Madhusudan;GATZEMEYER, John J.;JIMENEZ, Eduardo J.;KENNEDY, Sharon	1
37	COLGATE-PALMOLIVE COMPANY;WONG, Chi Shing;FONTANA, Jose Eder;FOCASSIO, Paulo	1
38	CRISP, Jackson	
39	CRISP, Jackson	1
40	DELTA OF SCIENCE APS;LYSTLUND, Thomas	
41	DELTA OF SCIENCE APS;LYSTLUND, Thomas	1
42	DENTALPOINT AG	
43	DENTALPOINT AG	1
44	DENTEK ORAL CARE INC	
45	DENTEK ORAL CARE INC	1
46	Dentsoll	
47	DENTSOLL KOREA CO., LTD.;KIM, Yun soon	1
48	DENTSOLL KOREA CO.,LTD.;KIM, Yun Soon	1
49	GC CORP;株式会社ジ シ	
50	GC CORP;株式会社ジ シ	1

Detailed Discussion – Landscaping Tasks

Technique

List Generation (Histograms)

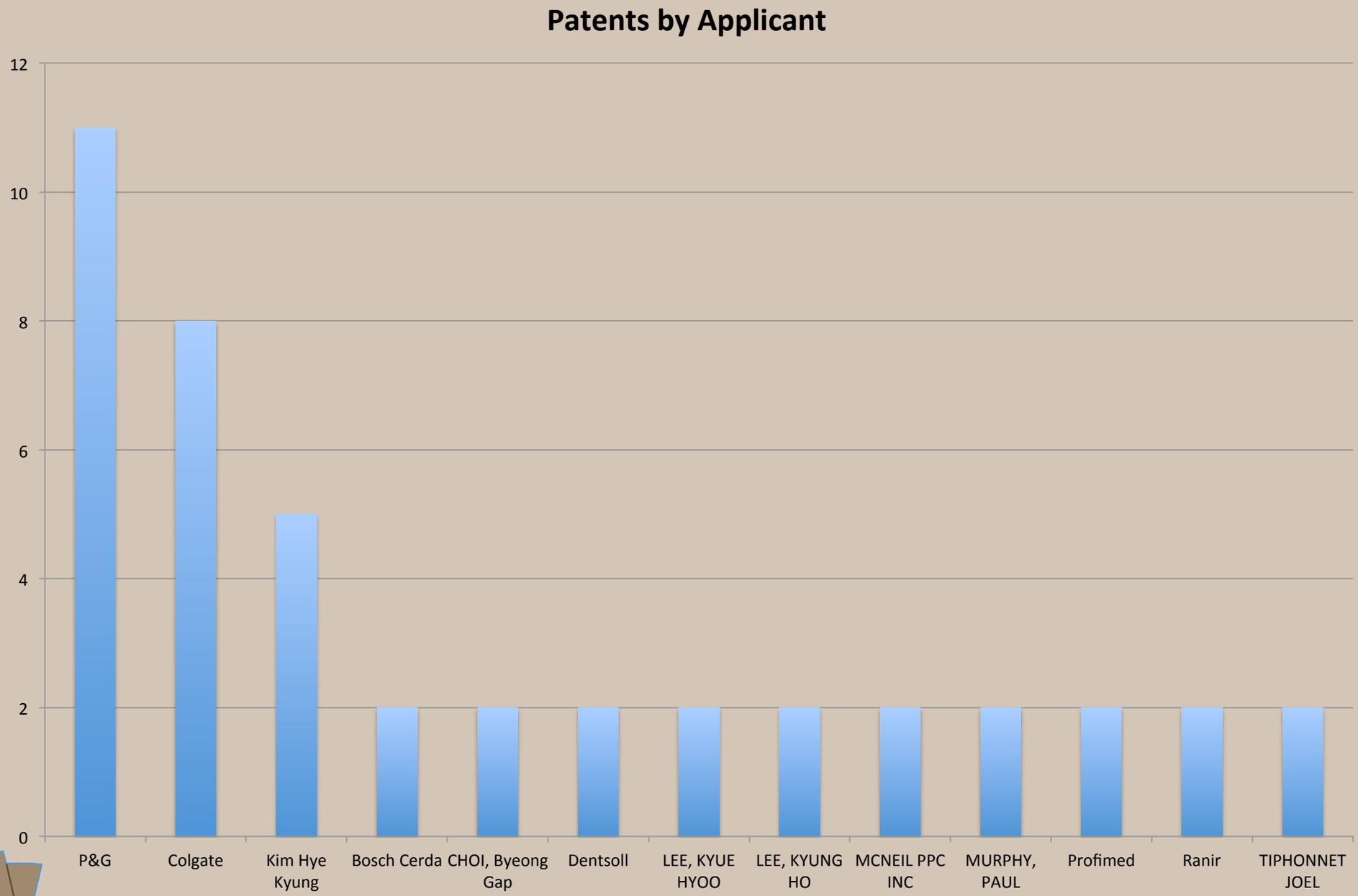
Definition

Provides counts of various patent related metrics within individual data fields

Utility

Allows the statistical comparison of two or more entities in the same data field

Histograms



Detailed Discussion – Landscaping Tasks

Technique

Co-Occurrence Matrices

Definition

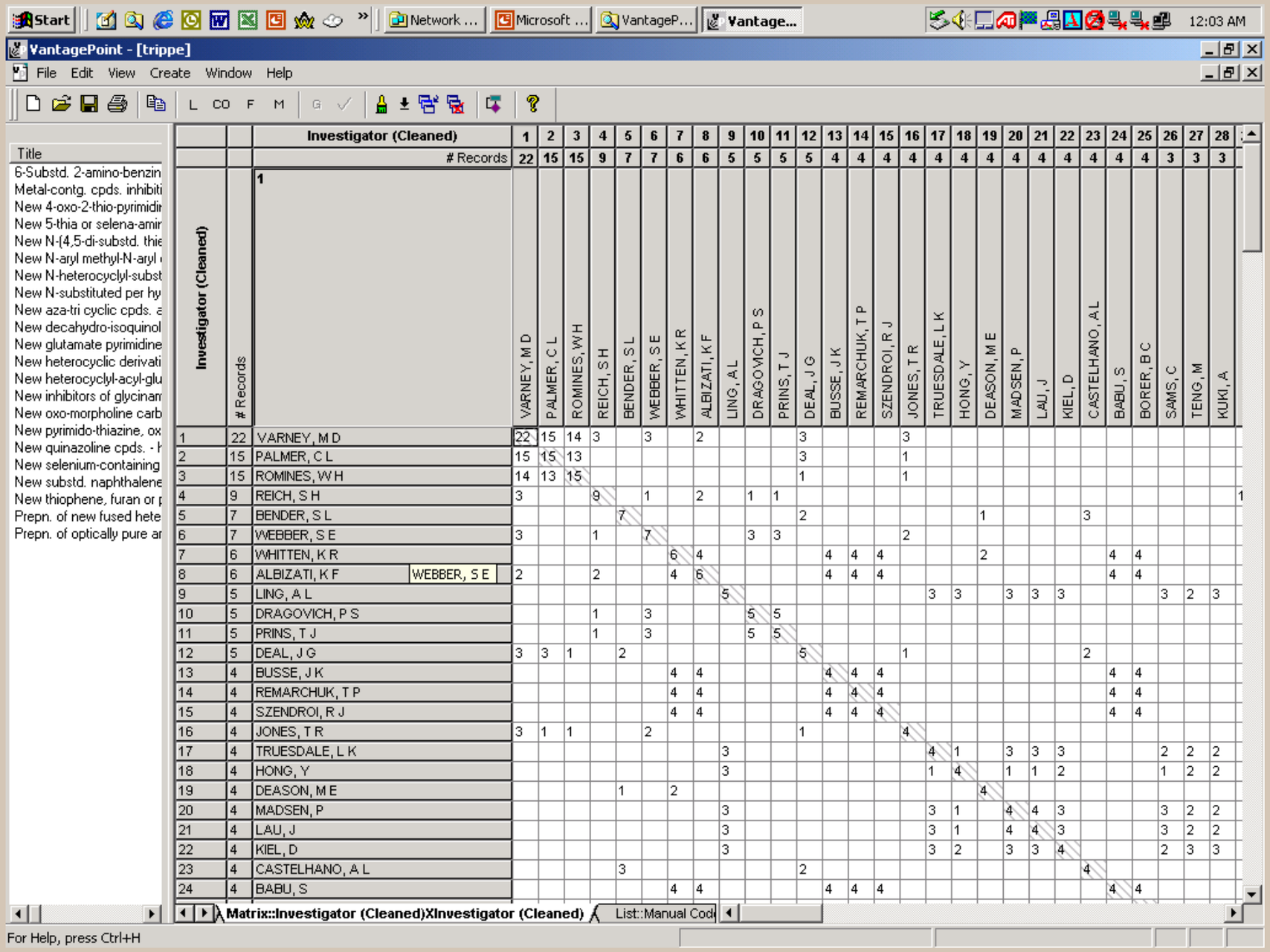
Data fields are placed on an X and Y-axis or on opposites sides of a circle. Number of overlapping occurrences of shared X and Y can be seen as numbers within the matrix

Utility

Allows connections to be made between two or more fields of information and provides a representation of how strong the connection is

Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Co-Occurrence Matrix



	Investigator (Cleaned)	# Records	VARNEY, M D	PALMER, C L	ROMINES, W H	REICH, S H	BENDER, S L	WEBBER, S E	WHITTEN, K R	ALBIZATI, K F	LING, A L	DRAGOVICH, P S	PRINS, T J	DEAL, J G	BUSSE, J K	REMARCHUK, T P	SZENDROI, R J	JONES, T R	TRUESDALE, L K	HONG, Y	DEASON, M E	MADSEN, P	LAU, J	KIEL, D	CASTELHANO, A L	BABU, S	BORER, B C	SAMS, C	TENG, M	KUKI, A
1	VARNEY, M D	22	22	15	14	3		3		2				3				3												
2	PALMER, C L	15	15	15	13									3				1												
3	ROMINES, W H	15	14	13	15									1				1												
4	REICH, S H	9	3			9		1		2		1	1																	1
5	BENDER, S L	7					7							2							1				3					
6	WEBBER, S E	7	3			1		7				3	3					2												
7	WHITTEN, K R	6							6	4					4	4	4				2					4	4			
8	ALBIZATI, K F WEBBER, S E	6	2			2			4	6					4	4	4									4	4			
9	LING, A L	5									5								3	3		3	3	3				3	2	3
10	DRAGOVICH, P S	5				1		3				5	5																	
11	PRINS, T J	5				1		3				5	5																	
12	DEAL, J G	5	3	3	1		2							5				1							2					
13	BUSSE, J K	4							4	4					4	4	4									4	4			
14	REMARCHUK, T P	4							4	4					4	4	4									4	4			
15	SZENDROI, R J	4							4	4					4	4	4									4	4			
16	JONES, T R	4	3	1	1			2						1				4												
17	TRUESDALE, L K	4									3								4	1		3	3	3				2	2	2
18	HONG, Y	4									3								1	4		1	1	2				1	2	2
19	DEASON, M E	4					1		2												4									
20	MADSEN, P	4									3								3	1		4	4	3				3	2	2
21	LAU, J	4									3								3	1		4	4	3				3	2	2
22	KIEL, D	4									3								3	2		3	3	4				2	3	3
23	CASTELHANO, A L	4					3							2											4					
24	BABU, S	4							4	4					4	4	4									4	4			

Detailed Discussion – Landscaping Tasks

Technique
Clustering of Data

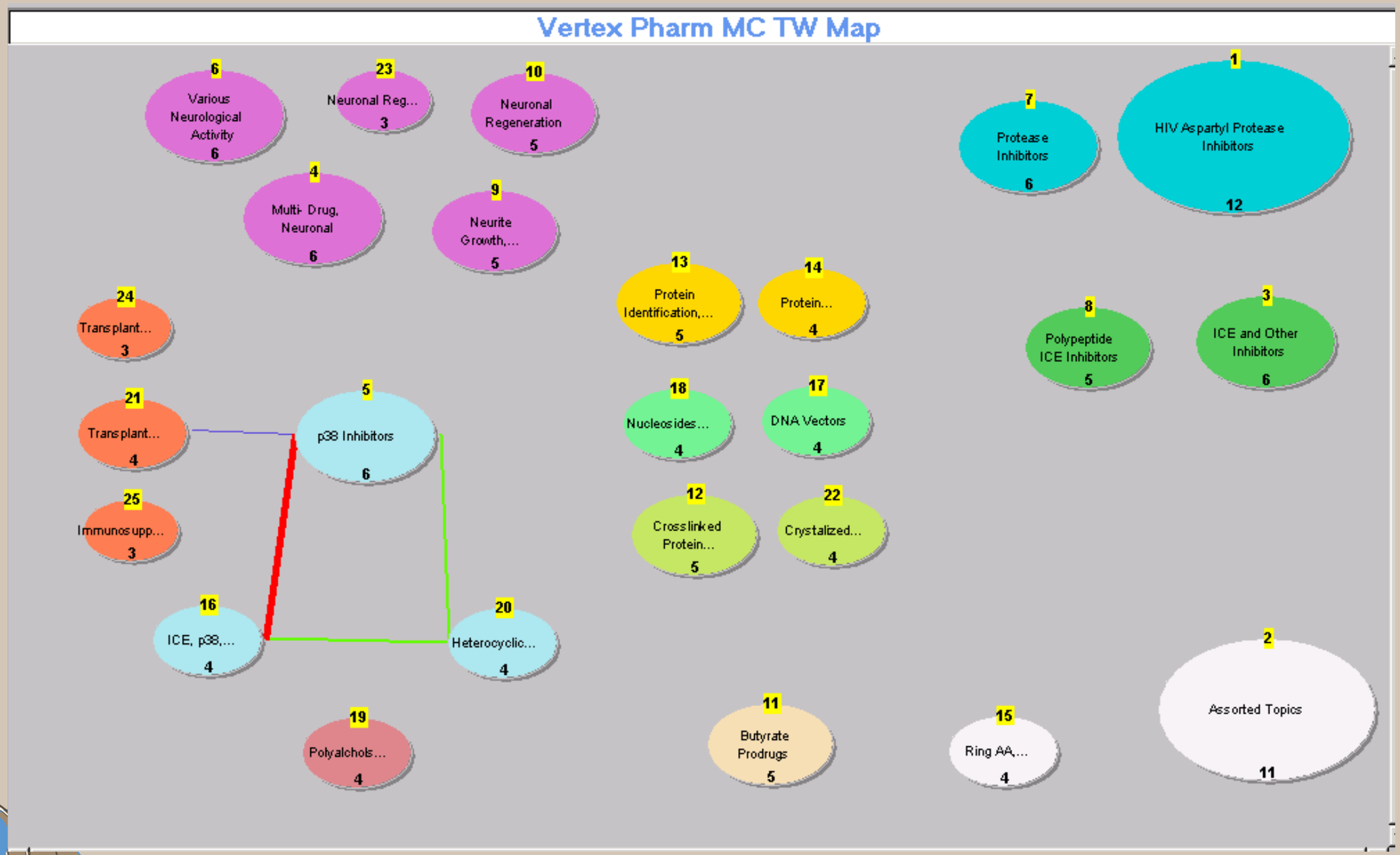
Definition

Associated with unsupervised methods of organizing document collections based on a similarity comparison between documents. With a fixed number of clusters identified at the outset, document collections that meet a threshold similarity component are grouped together.

Utility

Ideally, the documents within a cluster should be similar to one another but dissimilar to documents in the other clusters.

Clustering Fielded Data



Detailed Discussion – Landscaping Tasks

Technique

Classification of Data

Definition

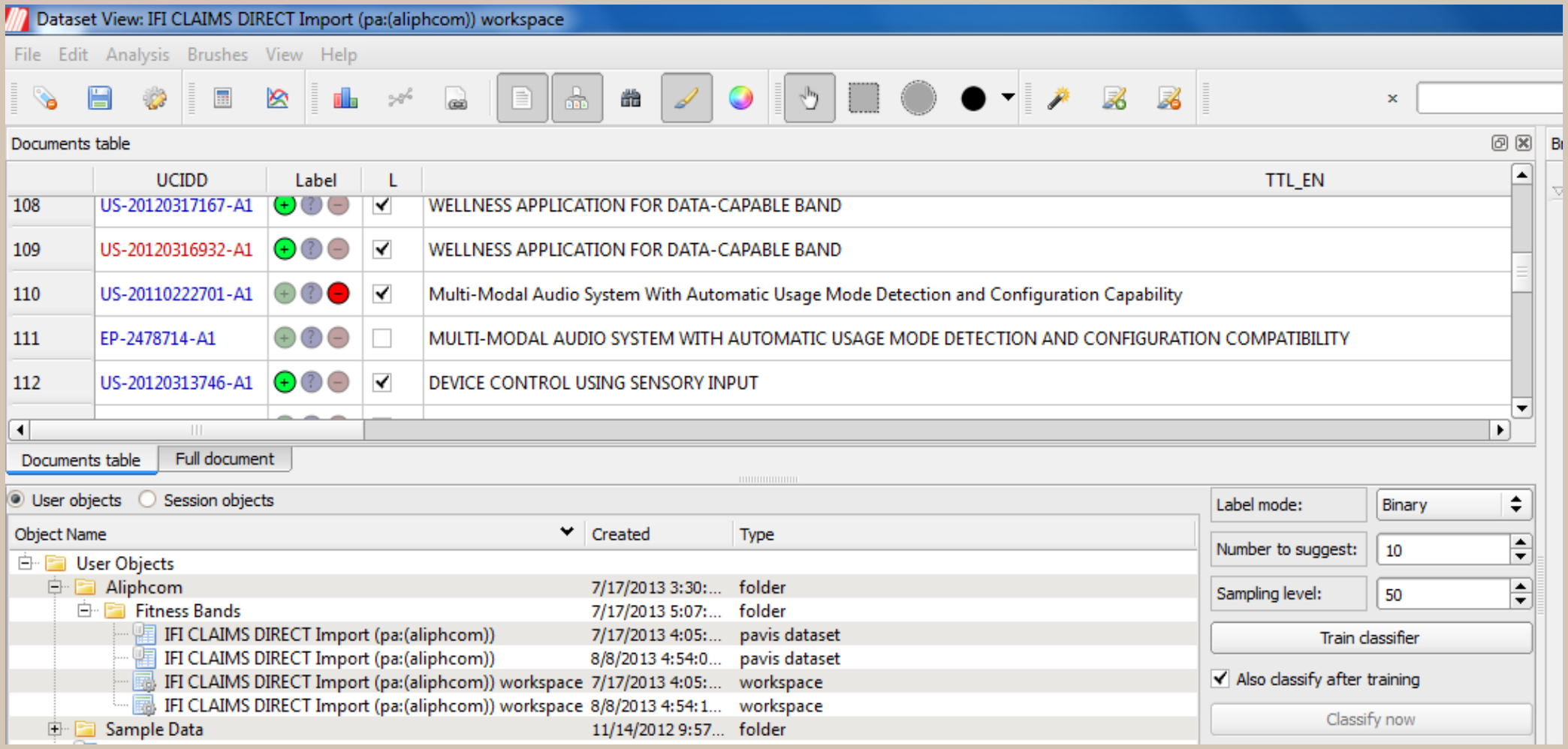
Accomplished using a supervised machine learning method that uses learning sets to identify key attributes of documents in a class. New documents are compared to the learning collections and assigned to a class based on their similarity to the documents that have already assigned to the class.

Utility

Can be used to categorize or prioritize search results

Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Binary Classification



Detailed Discussion – Landscaping Tasks

Technique

Spatial Concept Maps

Definition

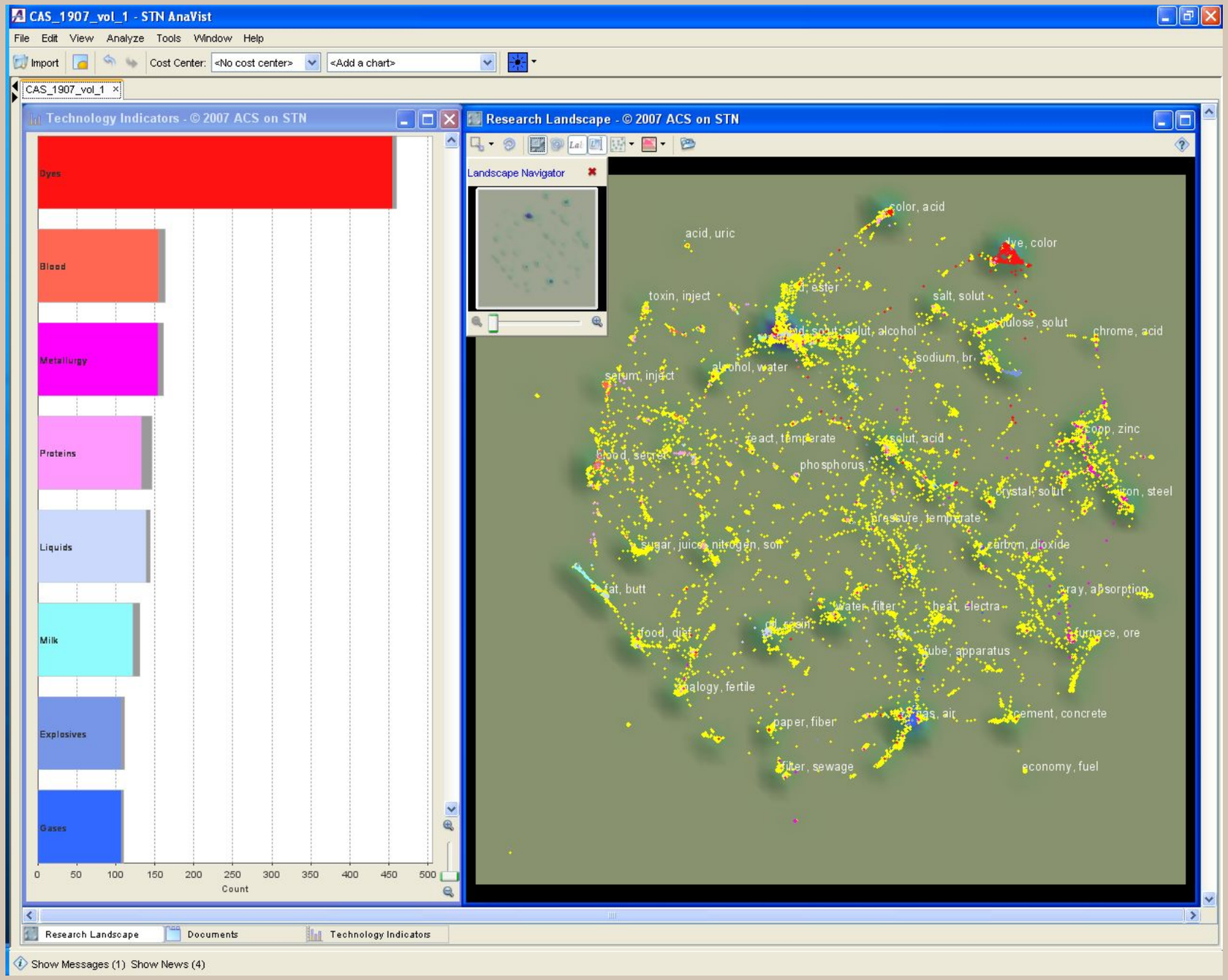
Document clusters are arranged in 2-dimensional space creating a map; Collections of documents, which share elements in common, are placed closer together geographically while collections with less similarity are placed further away

Utility

Allows relationships between clusters to be identified; Creates a visual representation of document collection at high-level view

Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Mapping Document Clusters



Detailed Discussion – Landscaping Tasks

Technique

Stacking or Layering Data

Definition

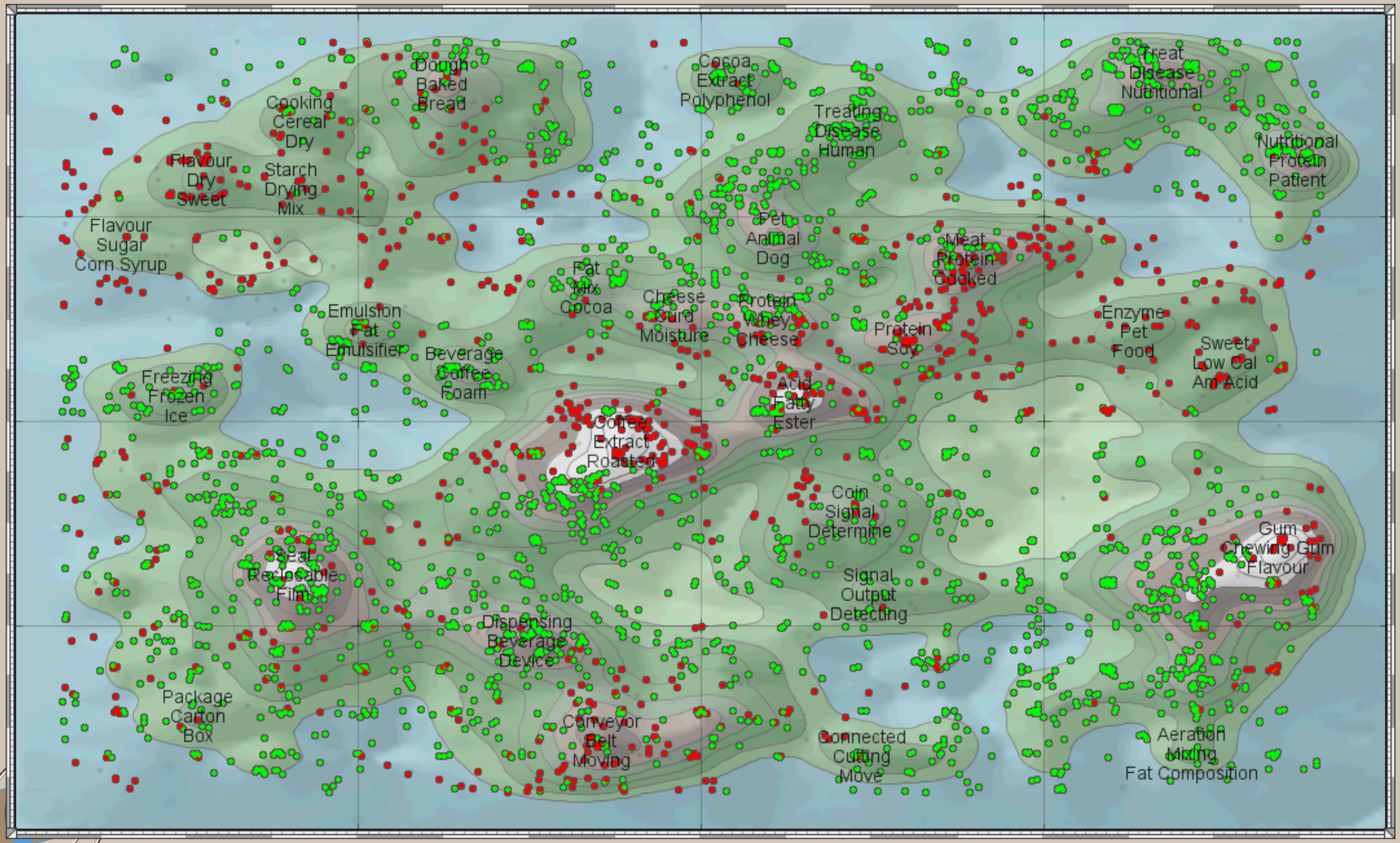
Analyses looking at a single variable, or field, can be inefficient and lack context. Positioning two types of visualizations next to one another, or adding overlays to an analysis, allows the analyst to reference several attributes of a data set simultaneously.

Utility

Additional context can be added without asking the client to refer back to previous illustrations.

Patinformatics, LLC®
Data Driven Decisions
Patent Strategy and Analytics Services

Adding Time Component to Map



Detailed Discussion – Landscaping Tasks

Technique

Network Analysis

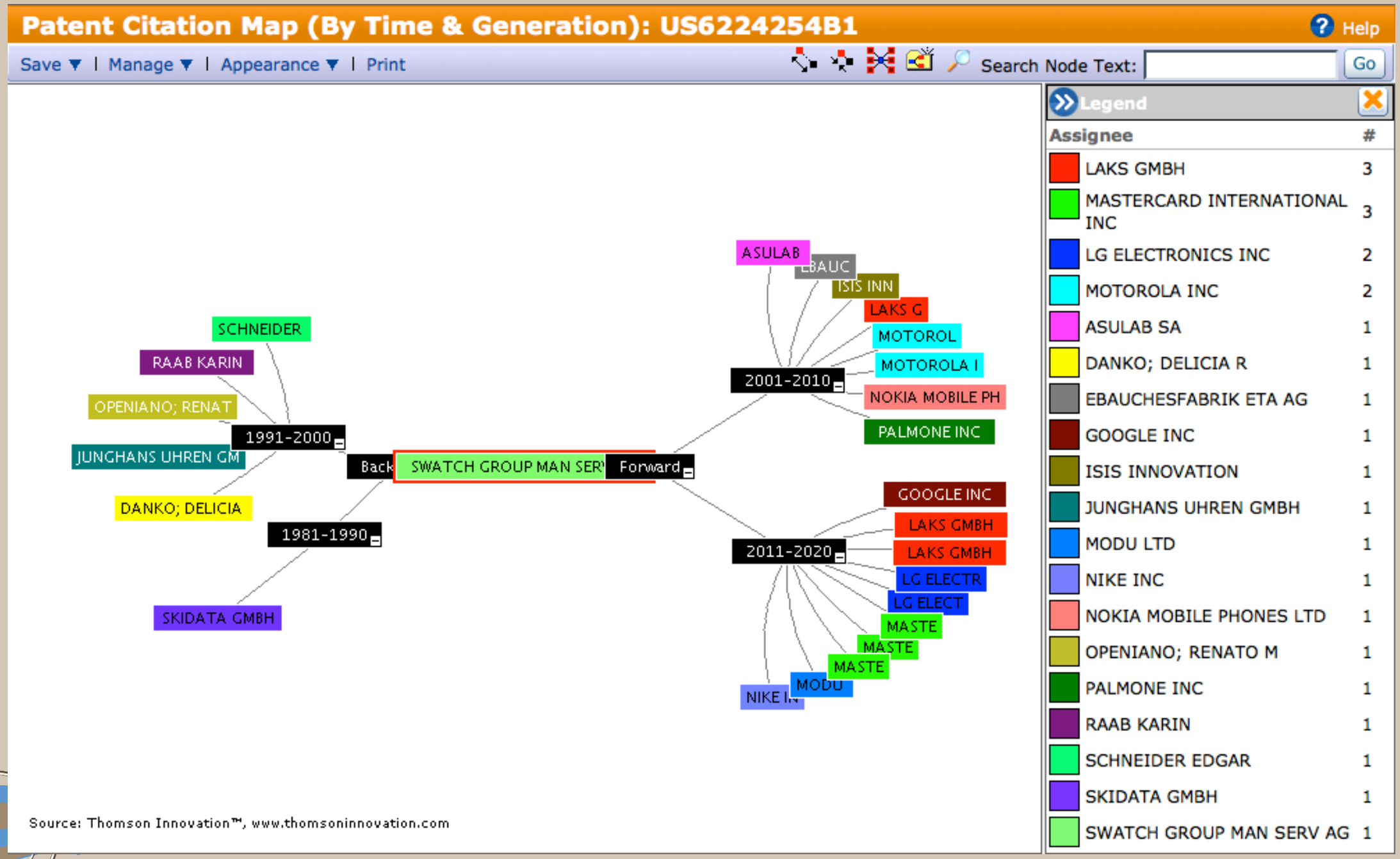
Definition

Network analysis is the viewing of relationships in terms of network theory , consisting of nodes, representing individual actors within the network, and ties, which represent relationships between the individuals, such as co-inventorship, co-assignment and co-citation.

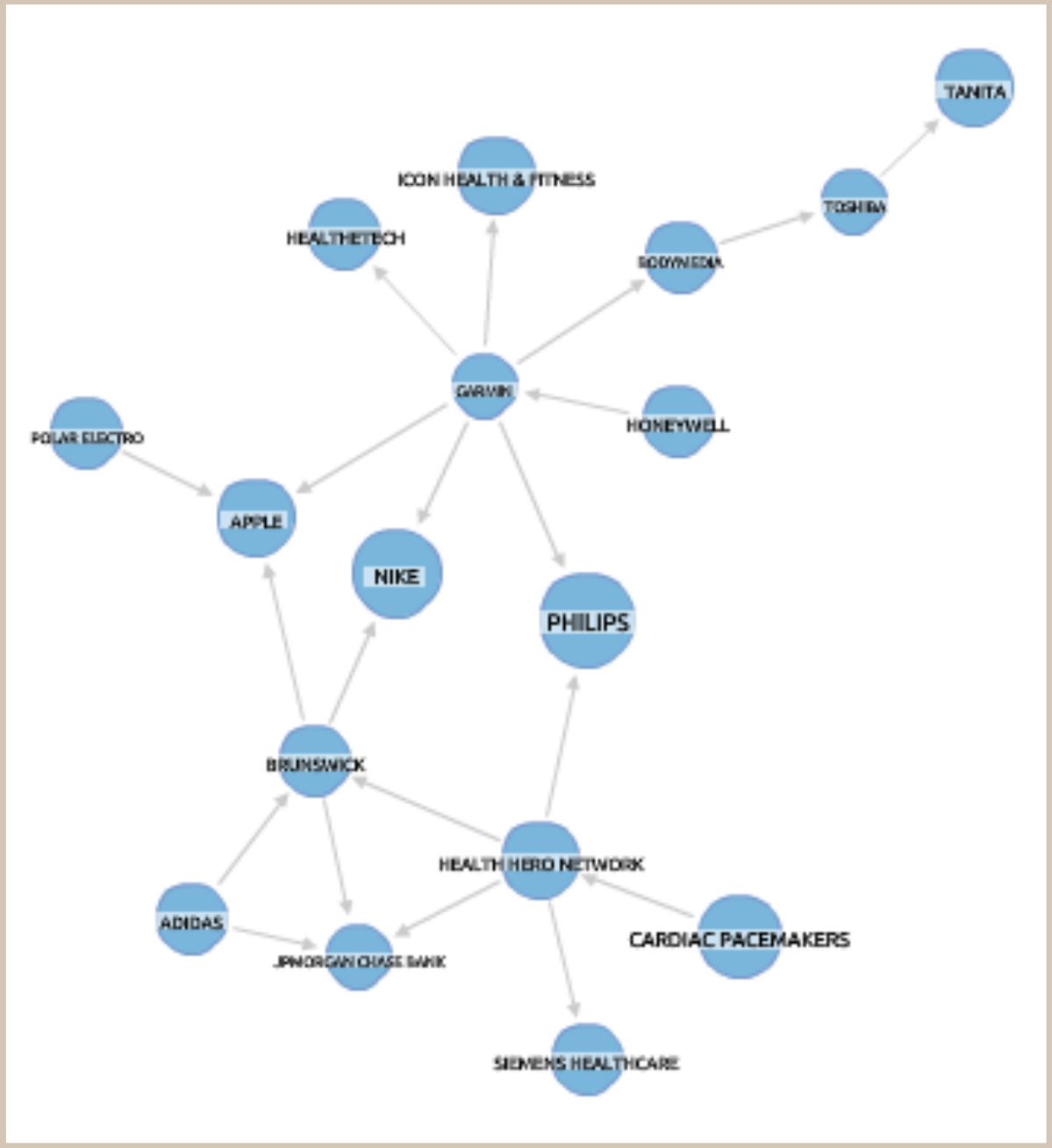
Utility

Hyperbolic trees are used to show relationships between patents that cite one another. Networks are built when more than one patent at a time is analyzed.

Citation Map – Hyperbolic Tree



Network Analysis



Detailed Discussion – Landscaping Tasks

Technique

Semantic Analysis

Definition

Subject/Action/Object (SAO) Functions

Parts of language that are used to describe the teachings that the author wants to portray. Key SAOs encapsulate the technical learnings contained in a document. SAOs can be described as problems and solutions.

Utility

By identifying SAOs the teachings of a document can be isolated and examined from the rest of the document creating a knowledge base

SAO Functions