PATINFORMATICS

DR. (MS.) SONALI CORREA

MSC., PH.D.

PATINFORMATICS

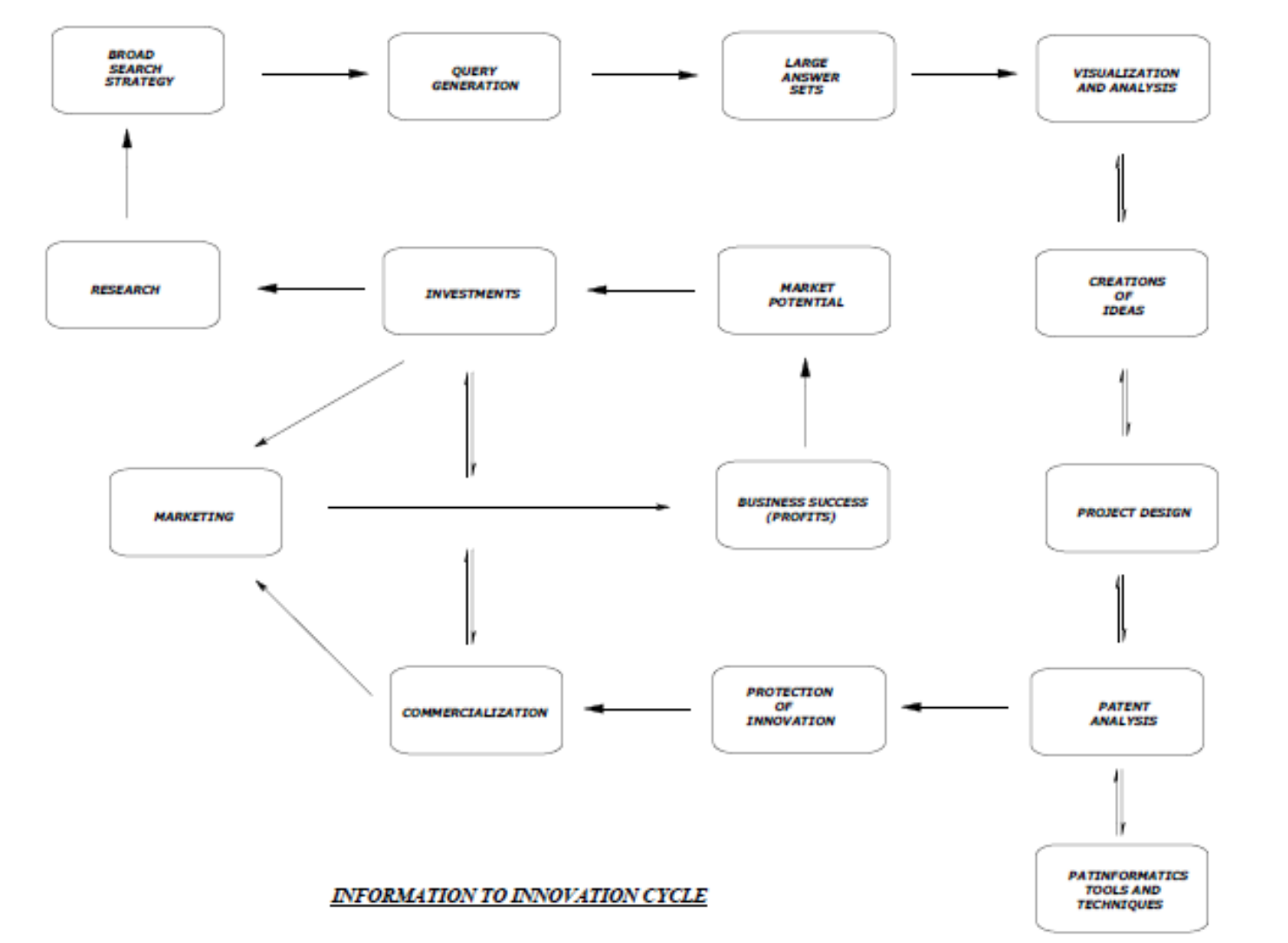
Introduction	• What is Patinformatics?
The Macro-Level forms	• patent intelligence, patent mapping, patent citation analysis
Patinformatics utility	
Available Patinformatics tool	

INTRODUCTION

- What is Patinformatics?
  - the science of analysing patent information to discover relationships and trends
  - Patinformatics is an emerging science which involves **analysis** of a set of **patent data** to **map relationships** and trends which would be otherwise difficult to do so when dealing with **patent documents on a one-to-one basis.**
  - The use of patent information and management has become a critical part in management of Research and Development (R&D), a tool for creative thinking, competitor's monitoring, technology assessment, new venture evaluation, input for licensing strategy, supporting mergers & acquisitions and human resource management.
  - Alternatively it may be defined as a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization and use of patent information
  - The use of patinformatics tools are in the updation, validation of technical and scientific information and their exploitation for the betterment of innovation.

Trippe, Anthony. (2003). Patinformatics: Tasks to tools. World Patent Information. 25. 211-221. 10.1016/S0172-2190(03)00079-6.

- Patinformatics is the judicious screening of information sources to transform vast amount of data into specific information and specific information into knowledge for the intended purpose of making better and faster decisions in various fields of science, technology andbusiness. Patinformatics may also be defined as an application of informatics method to solve patent related problems
- The first ever Indian conference on Patinformatics, titled ‘Recent Trends in Patinformatics’, was held from the 9th to 12th of December,2009 at the National Chemical Laboratory, Pune to mark the 26th National Convention of SIS (Society for Information Science) which was co hosted by URDIP (Unit for Research and Development of Information Products).
- The conference was aimed at answering several pertinent questions about the need, method and application of patent analysis as well as depicting both the Indian and Global status in this field.



INFORMATION TO INNOVATION CYCLE

- organization to start or to initiate a project:
- terms of player, relevant art, first hand information about the field
- knowledge of developed techniques of patinformatics stops overlapping R&D work and projects and thereby prevents unnecessary investments.
- The various techniques of patinformatics such as integration of patent data by importing the patent data from commercial and non-commercial databases in the various formats including XML/CSV/RTF etc or exporting it in various formats, capturing claims section separately in any of the customizable format or generating family equivalents portfolio by using patinformatics tools helps in collecting the maximum patent information.

PATENT ANALYSIS

- Patent information is the information found in patent documents and the associated databases which includes technical, legal and business-related informations (80% of this information is found in patent documents)
- The patent information plays a key role starting from an idea of an innovation to the fruitful output in the form of commercialization to give economic benefits to the organization and ultimately to progress..
- Patent analysis (or patent analytics) is a policy and decision-making support tool which uses information included in patent documents and other information from the patent lifecycle (patent information) to identify technology and patenting trends, and is a major form of Intellectual property
- Patent analysis consists of the evaluation and understanding of the innovation and the development of various new trends of the technologies andheir competitive positioning.
- The different techniques of patinformatics useful for analyses are data cleanup and grouping of concepts, clustering of structured and unstructured data, mapping, citation analysis, data mining, visualization, list generation, co-occurency matrices & circle graphs and SAO functions etc.

MACRO- LEVEL FORMS OF PATENT INFORMATION ANALYSIS

Patent intelligence

- The use of patent information to identify the technical capabilities of an organization and the use of that intelligence to develop a strategy for strategic technical planning

Patent mapping

- Sometimes described as white space mapping, which uses published patent data to create a graphical or physical representation of the relevant art pertaining to a particular subject area or novel invention

Patent citation analysis

- the study of patent citations for potentially determining a patent's value or, perhaps more reliably, the identification of potential licensing partners or leads based on the citation of an organization's patents by another company in the same or a completely different market space

Patinformatics can also cover additional applications of patent information involving a subsequent analysis step. The key underlying property in each of these diverse areas is the analysis step.

PATINFORMATIC PRINCIPLES

- **Data mining** involves the **extraction of fielded data** and its analysis.
- This means analyzing the bibliographic information contained within patents.
- Example, someone might want to examine the relationship between patent assignees and International Patent Classification (IPC) codes for a specific area of technology. Mining or mapping this information can give an idea of the major players in a technology area and what type of work they generally focus on.
- Using **Derwent data**, a similar analysis could replace IPC codes with Derwent manual codes.
- Using advanced data science algorithms, Derwent Data Analyzer™ is a desktop data-mining platform that converts patent data, scientific literature and your own business intelligence into actionable, commercial insight.

https://clarivate.com/products/ip-intelligence/patent-intelligence-software/derwent-data-analyzer/

- **Text mining or mapping** typically involves clustering or categorizing documents based on the major concepts contained.
- The data source is unstructured text, it is not fielded, and the only structure within the material comes from what the author applied when writing the document and how they built relationships between different concepts and ideas.
- For example, you could collect patents from a specific patent assignee and analyze the text of those documents. In a cluster map, the software would extract the major concepts found and create clusters of documents concept by concept. The software would then visualize these clusters in some fashion, creating a map. By looking at the clusters created (and subsequently, the documents themselves, but now with an organized method), you can quickly get a general idea of the concepts that this organization is working on and how these concepts interrelate.

LINEAR ANALYSIS OF PATENT INFORMATION

- – Create a tool kit of patinformatics tools
  - patinformatics can include patent mapping, citation analysis, co-occurrency analysis, thematic mapping, temporal visualization, and various other techniques
- – Understand the Need Behind the Need
  - In patinformatics it is absolutely essential that the business need for intelligence is clearly understood before anything else begins. Analysts need to understand how the data will be used and who will use it.
  - They need to know what type of story to tell in order to represent their intelligence work in such a way that the person receiving it will understand it and will stand the greatest chance of putting it into business practice.
- – The Need Drives the Question
  - In a true linear sense, once the need is understood, then the analyst and client can work together to formulate questions to supply intelligence that will impact the underlying business decision.
  - For example, a business may need to gain additional insight on how the research and development progress works for a particular company (say Company Y again).
  - Questions: understanding what research and development projects Company Y conducts in its 10 different research facilities in the U.S. becomes an interesting question.
  - Additional questions might include the following: Where do the inventors on their U.S. patents live? What patenting topics are closer to basic science? Which apply more to process technologies? By asking a number of compelling questions and compiling intelligence on each of them, an analyst can begin to paint a mosaic of the dynamics associated with the business need. Examining all the dynamics will lead the analyst to draw conclusions.

- – The Question Drives the Data
  - Referred to as the scientific method, this process involves the formation of a hypothesis, experimentation to determine the validity of the hypothesis, and verification of the validity of the experimentation and of the conclusions drawn based on experimental results. In the realm of patinformatics, the gathering of data is directly analogous to the idea of preparing an experiment to support or dispute a hypothesis. Selection of the appropriate tool is also important to the process.
- – The Data Drives the Tool
  - Requirement of very specific types of data. ‘
  - In these circumstances, the tool selected must not **only allow for the analysis necessary** to provide the insight but must also **work with the most appreciate data source**
  - Example: the inventors on Company Y's U.S. patents live, then the data will have to include the inventor's address information, which appears on the front page of all U.S. patents. Perhaps more importantly, this data must be available in an electronic format for importing into the appropriate analysis tool
  - Another principle is the idea of Actionable Intelligence. This idea dictates that intelligence is only useful if it is applied to a business question and more importantly used to make a business decision.

PATENT MAPPING

- Patent landscape analysis, often referred to as “patent mapping” is a proven multistep process, employing computer software and human intelligence, to parse through, organize and extract value from this vast amount of information.
- provides insight into the innovations that underlie technology and products
- A completed patent landscape analysis project consists of a set of technical references and accompanying analytics from which important legal, business, and technology information can be extracted.

Benefits

- This information enables large corporations, startups, universities, research institutions, and investors to understand and make informed decisions prior to investing time and money into new technology and product development opportunities.

Application

- The information derived from a patent landscape analysis is applied within an organization to generate novel technology, to identify possible companies or technologies to license or acquire, to design around others’ technology to side-step litigation, or to avoid spending time and money on duplicating technology that has been developed previously and may already be on the market.

BENEFITS & BENEFICIARIES

- **Researchers and scientists gain current awareness** (understanding the "state of the art") and learn important details about a given technology and potentially build, expand, improve, or modify that art.
- **Business development teams identify who are the important assignees (owners) of the relevant technology**. This is useful for considering potential merger/acquisition partners, in/out licensing opportunities (open innovation), competitors, suppliers, or customers and to assess whether it makes more financial sense to develop the technology in-house.
- **Marketing, competitive intelligence, commercial strategy, and human resource teams gain early insight into what innovations competitors and their inventors have in the development pipeline** and what products competitors are betting will be successful in the market. Looking at patent filing trends in conjunction with market research is useful to determine the commercial phase of the technology and whether it is in its infancy or is in a mature or declining phase.
- Legal teams, whether in-house our outside counsel, use a patent landscape analysis to understand the relationship between competitive products and patent protection for competitors and where potential product or process innovations may be barred because of numerous patent references or broad patent claim language. If performed early in an R&D project, answering this question can save considerable time, money, and human resources.
- Executives, high-level strategists, and investors gain insight into the opportunities and risks that exist in light of competitive R&D and product development initiatives, new "areas" for acquisition or investment, and how to develop and maintain a sustainable growth strategy.
- Human resources and recruiters can identify and hire the leading innovators (inventors) in a technology.

- Here are some common applications of patent analysis:
- 1. **Competitor analysis**: Analyze your competitors' patent portfolios to identify areas of strength and weakness, and to anticipate potential future developments.
- 2. **Innovation tracking**: Monitor patents related to emerging technologies or trends to identify early-stage innovations and potential game-changers.
- 3. **IP portfolio management**: Use data analytics to optimize your own patent portfolio, identifying areas for consolidation, expansion, or divestment.
- 4. **Licensing and royalties**: Analyze patent data to identify potential licensing opportunities and negotiate more favorable royalty agreements.
- 5. **Research and development**: Use patent analysis to identify areas where R&D investment may be most effective, and to inform product development decisions.
- 6. **Market research**: Analyze patents related to specific industries or markets to gain insights into market trends, gaps, and opportunities.

- Some common techniques used in patent analysis include:
- 1. **Text mining**: Extract relevant information from patent text, such as keywords, concepts, and entities.
- 2. **Network analysis**: Analyze the relationships between patents, inventors, companies, and technologies.
- 3. **Clustering**: Group patents based on similarities in their content, classification, or other characteristics.
- 4. **Classification**: Assign patents to categories or classes based on their content or technical fields.
- 5. **Machine learning**: Use machine learning algorithms to predict the likelihood of patent infringement, identify potential litigation risks, or forecast future innovation trends.
- 6. **Data visualization**: Use visualizations to represent complex patent data in a clear and intuitive way.
- Some popular tools and platforms for patent analysis include:
- 1. **Patent databases**: Commercial databases like PatentSight, Patinformatics, and Questel's Total Intellectual Property Solutions (TIPS) provide access to large collections of patent data.
- 2. **Patent analysis software**: Tools like PatentScope, PatentOptimizer, and PatentSight's Analytics Platform offer specialized features for analyzing and visualizing patent data.
- 3. **Open-source tools**: Tools like OpenPatentSearch, PatentDB, and PatentGraph provide free access to patent data and allow users to develop their own custom analysis tools.

- Some common challenges in patent analysis include:
- 1. **Data quality issues**: Patent data can be incomplete, inconsistent, or difficult to interpret.
- 2. **Scalability**: Working with large collections of patent data can be computationally intensive and require significant resources.
- 3. **Complexity**: Patent analysis often involves complex technical concepts and requires specialized knowledge.
- 4. **Interpretation**: Extracting meaningful insights from patent data requires careful interpretation and critical thinking.
- Overall, patent analysis is a powerful tool for gaining insights into the competitive landscape

PERFORMING PATENT LANDSCAPE ANALYSIS

- A patent landscape analysis is a type of patent analysis that provides a comprehensive overview of the patent landscape in a specific technology or market. This type of analysis is used to identify trends, patterns, and relationships in the patent data, and to gain insights into the competitive landscape.
- Here are the general steps involved in performing a patent landscape analysis:
- Define the scope: Define the technology or market that you want to analyze, and identify the relevant patents and patent applications.
- Collect patent data: Collect the relevant patent data from various sources, such as patent databases, online patent portals, and patent offices.
- Clean and preprocess the data: Clean and preprocess the data to ensure that it is consistent and accurate.
- Analyze the data: Use various techniques, such as text mining, clustering, and network analysis, to analyze the patent data and identify trends, patterns, and relationships.
- Visualize the results: Use visualization tools to present the results in a clear and intuitive way.

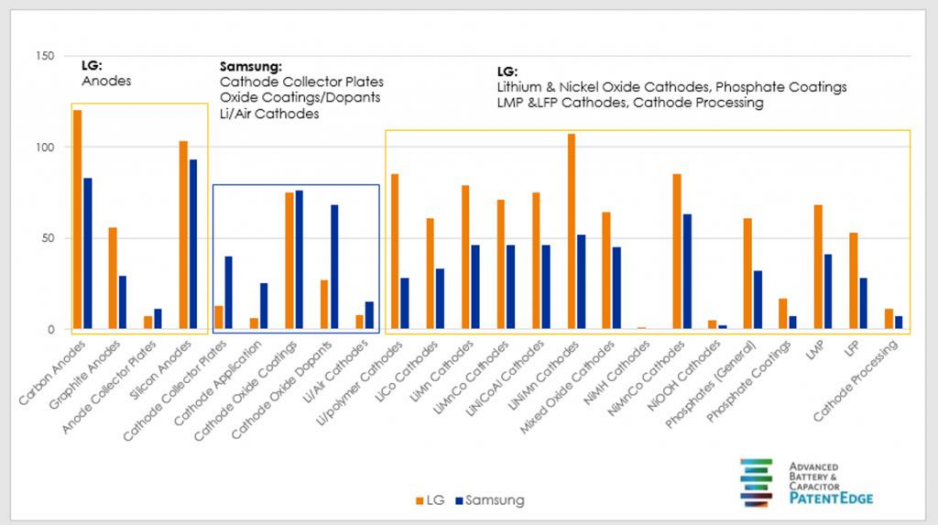
- Some common techniques used in patent landscape analysis include:
- Text mining: Extract relevant information from patent text, such as keywords, concepts, and entities.
- Clustering: Group patents based on similarities in their content, classification, or other characteristics.
- Network analysis: Analyze the relationships between patents, inventors, companies, and technologies.
- Classification: Assign patents to categories or classes based on their content or technical fields.
- Machine learning: Use machine learning algorithms to predict the likelihood of patent infringement, identify potential litigation risks, or forecast future innovation trends.

- Some common tools and platforms used in patent landscape analysis include:

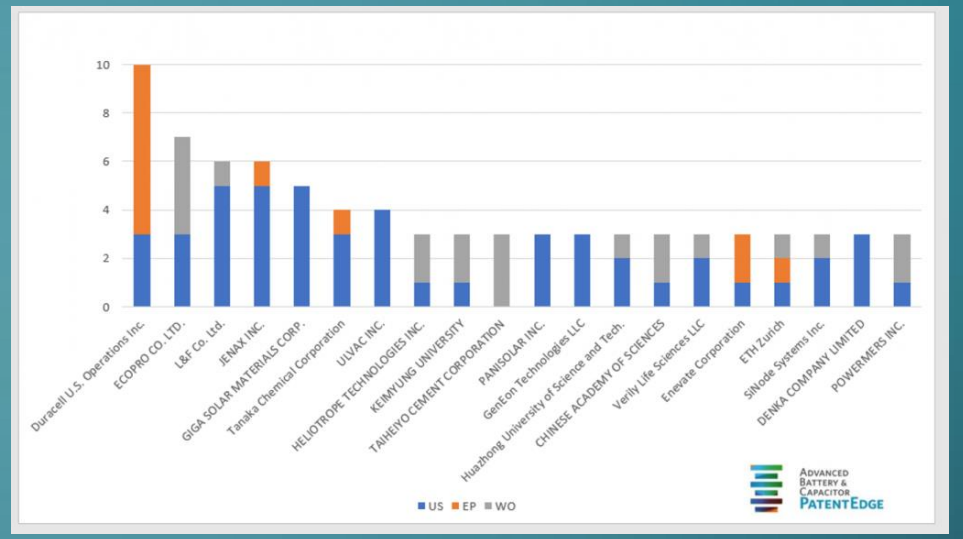
- Patent databases: Commercial databases like PatentSight, Patinformatics, and Questel's Total Intellectual Property Solutions (TIPS) provide access to large collections of patent data.
- Patent analysis software: Tools like PatentScope, PatentOptimizer, and PatentSight's Analytics Platform offer specialized features for analyzing and visualizing patent data.
- Open-source tools: Tools like OpenPatentSearch, PatentDB, and PatentGraph provide free access to patent data and allow users to develop their own custom analysis tools.

ANALYZE THE DATA

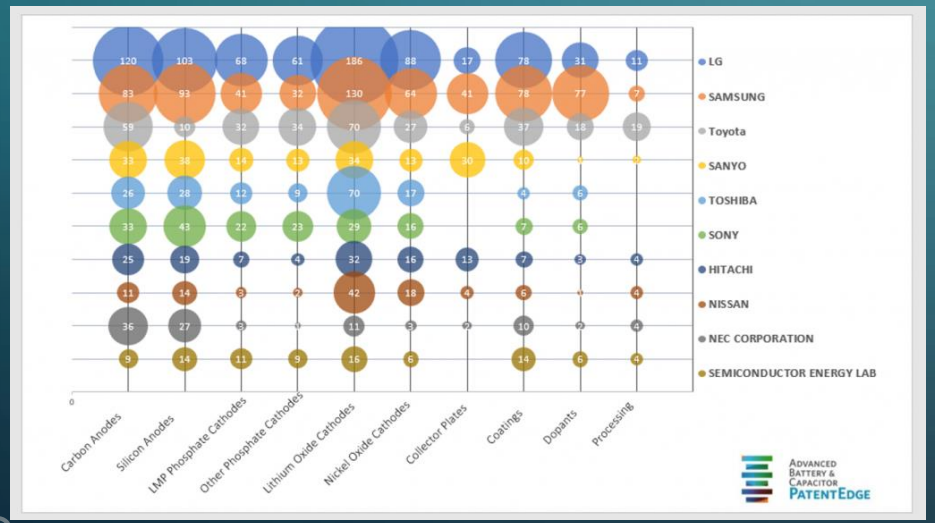
- **Patent filing velocity charts enable you to visualize the amount of change in activity in each technology, product, or application area over time**. This is useful to understand how much growth has occurred in the area, and be predictive about when activity might start to level off. This can be particularly useful if paired with market research including related product sales data.
- Top assignee tables provide insight into which large companies, startups, universities, and others are most focused in each technology, product, or application area. This is useful to identify potential partners, customers, licensees, acquisition candidates or organizations that might be using your technology evidenced by information that is disclosed in their patents (i.e. those discussing technology in their patent filings to which you own the underlying patent rights).
- Combined is used to visualize which players are spending money (via patent filings) in which categories more recently and which companies may have slowed investment in a technology area. This can further lead to comparing the R&D spending (via patent filing activity) in a variety of technology areas compared to other competitors R&D spending in the same and/or other technology, product, or application areas.
- Patent claims in each technology, product, or market area can be scrutinized and mapped against other patents or products for legal analysis pertaining to validity, patentability, or freedom to operate.
- High-level patent landscape maps or visualizations can be analyzed for research, discovery and “white space” opportunities. These graphic visualizations can also be useful for communication and marketing tools – especially to non-patent experts and investors.



Depicts the worldwide patent activity of LG and Samsung from 2016 in the area of battery electrodes. Chart courtesy of The Advanced Battery & Capacitor PatentEdge.



Depicts the worldwide patent activity of 2016-new entrants in battery electrodes, (those company's that did not have a patent publish related to Electrodes in the prior year). Chart courtesy of The Advanced Battery & Capacitor PatentEdge.



Depicts the top assignees patent activity of 2016 as relates to the top 10 electrodes categories from the ABC PatentEdge. Chart courtesy of The Advanced Battery & Capacitor PatentEdge.

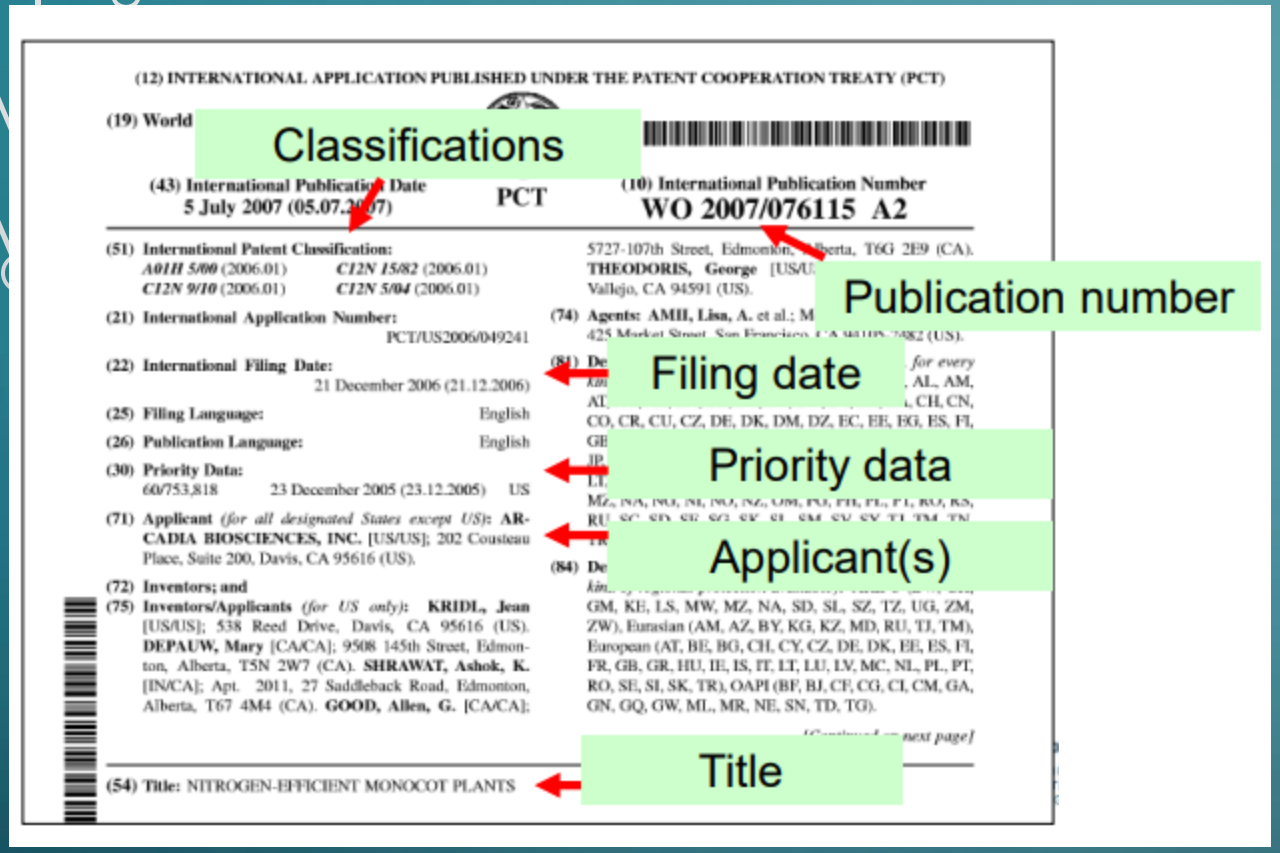
PATENT INTELLIGENCE

PATINFORMATIC TASK

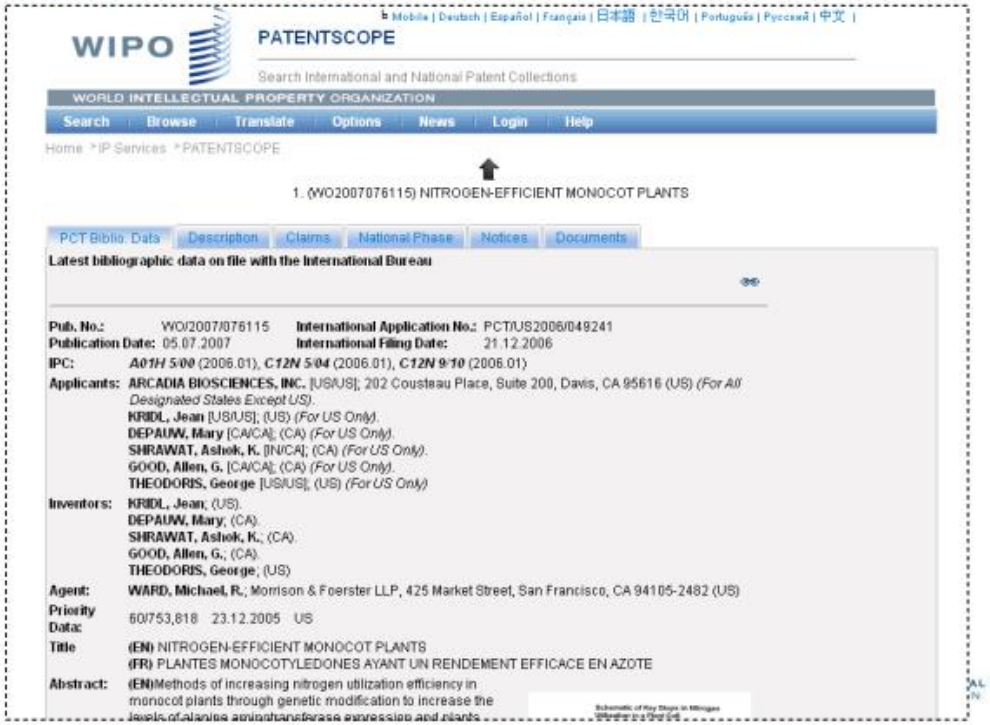
Summary of patinformatics tasks, definitions, utility and tools

Technique	Patinformatics definition	Patinformatics utility	Available Patinformatics tool
List Cleanup and Grouping of Concepts	Manual or automatic standardization of terms within a data field	List cleanup is required in order to produce statistically relevant results. Grouping allows synonymous terms to be combined together so that their true value in a data set can be accurately assessed	VantagePoint, MS Excel, ClearResearch, OmniViz, Aureka ThemeScape
List Generation (Histograms)	Provides counts of various patent related metrics within individual data fields	Allows the statistical comparison of two or more entities in the same data field	VantagePoint, MS Excel, ClearResearch, Aureka Reporting, Knowledgist, Technology Watch, Wisdomain Analysis Module, Delphion PatLab II, SciFinder
Co-occurrency Matrices and Circle Graphs	Data fields are placed on an *X* and *Y* axes or on opposites sides of a circle. Number of overlapping occurrences of shared *X* and *Y* can be seen as numbers within the matrix or as lines of increasing width connecting items on the circle	Allows connections to be made between two or more fields of information and provides a representation of how strong the connection is	SciFinder Panorama, VantagePoint, ClearResearch, Aureka Reporting, Wisdomain Analysis Module, Delphion PatLab II
Clustering of Structured (Fielded) Data	Intellectually assigned classification systems produce a standardized code that can be used as a means of organizing documents that share a similar coding structure	Documents, which share a high percentage of codes in common, are likely to be similar. Allows a large number of documents to be organized	Technology Watch, ClearResearch, OmniViz, VantagePoint

Clustering of Unstructured (Text) Data	Raw text is processed to identify concepts and phrases contained within. As with the clustering of structured data, concepts instead of codes are used to group documents that share a high degree of overlap	Documents, which share a high percentage of concepts in common, are likely to be similar. Allows a large number of documents to be organized	Aureka ThemeScape, ClearResearch, OmniViz, Vivisimo, Delphion Text Clustering, VantagePoint
Mapping Document Clusters	Document clusters are arranged in two-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	Allows relationships between clusters to be identified. Creates a visual representation of a document collection from a high-level view	Aureka ThemeScape, Technology Watch, ClearResearch, OmniViz, VantagePoint
Adding Temporal Component to Cluster Map	A time dimension can be called out on a map usually by means of alternate colors	User can follow the progression of a subject as it develops or evolves	Aureka ThemeScape, ClearResearch, OmniViz
Citation Analysis	When patent documents are examined relevant prior art is mentioned on the search report or on the front page of the documents. The number of citations can be counted or followed as they link documents together	Hyperbolic trees are used to show relationships between patents that cite one another. Citation counts are used to discover potentially pivotal documents	M-CAM, Aureka Citation Trees, Delphion Citation Link, Metrics Group Citation Bridge, Wisdomain Citation Module
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. By identifying SAOs the teachings of a document can be isolated and examined from the rest of the document creating a knowledge base	Knowledgist



Structured/fielded data



Structure/ Fielded data

- **Visualization** is an especially effective way of representing the results of patent analysis.
  - Visualization methods are considered to be proper for representing patent information and its analysis results. Therefore, patent visualization is a key tool for patent analysis and detection of technical opportunities. Visual analysis is a powerful method to address challenges posed by patent information overload. An increasingly used mechanism is visual analysis of patents that involves 2-dimensional spatial patent visualization and leverages the capability of the human visual system to identify patterns and anomalies.
  - key advantages drastically reduce the time-to-insights and explore intellectual property-congested technology spaces in a swift but efficient fashion
  - Visual patent analysis is used to explore the relationship between companies, inventors and their research and to explore semantic relationships between patent content.
- **Co-occurrence** links between entities are usually calculated using matrix multiplication.
  - It allows connections to be made between two or more fields of information and provides a representation of how strong the connection is. In the case of non-binary occurrence matrices, it is necessary to use a generalized matrix multiplication method to calculate co-occurrence counts.

PATENT DATABASES

- A database is simply an organized collection of related data, typically stored on disk/server, and accessible by possibly many concurrent users. The databases are needed for collection and preservation of data, making data easy to find and search, standardizing data representation and finally organizing data into knowledge. A database is a collection of non-redundant data which can be shared by different application systems.
- LecturePatent_22\Lecture_Patinformatics\SSRN-id1566067.pdf

Patinformatics tools and their vendors

Tool name	Vendor site URL
Aureka Citation Tree	http://www.aurigin.com/aureka_online.html#citation
Aureka Reporting Module	http://www.aurigin.com/aureka_online.html#reporting
Aureka ThemeScape	http://www.aurigin.com/aureka_online.html#themescape
Autonomy	http://www.autonomy.com
Citation Bridge—Metrics Group	http://www.patentcitations.com
Citation Link—Delphion	http://www.delphion.com/products/research/products-citelink
ClearLab Studio—ClearForest	http://www.clearforest.com
ClearResearch—ClearForest	http://www.clearforest.com/products/products.asp?id=1
Delphion PatLab II	http://www.delphion.com/products/research/products-patlab
Delphion Text Clustering	http://www.delphion.com/products/research/products-cluster
DOORS—M-Cam	http://www.m-cam.com/doors/
InXight Categorizer	http://www.inxight.com/products/categorizer/
Knowledgeist—Invention Machine Corp.	http://www.invention-machine.com/prodserv/knowledgist.cfm
OmniViz	http://www.omniviz.com
SciFinder Panorama—Chemical Abstracts	http://www.cas.org/SCIFINDER/panorama.html
SEMIO Map	http://www.entrieva.com/entrieva/products/semiomap.asp?Hdr=semiomap
Technology Watch—IBM Synthema	http://www.synthema.it
VantagePoint—Search Technologies	http://www.thevantagepoint.com
Vivisimo	http://www.vivisimo.com
Wisdomain Analysis Module	http://www.wisdomain.com/AnalysisModule.htm
Wisdomain Citation Module	http://www.wisdomain.com/CitationModule.htm

Matrix of patinformatics tasks and tools

Tool	List Cleanup and Grouping of Concepts	List Generation	Co-occurrency Matrices and Circle Graphs	Clustering of Structured (Fielded) Data	Clustering of Unstructured (Text) Data	Mapping Document Clusters	Adding Temporal Component to Cluster Map	Citation Analysis	SAO Functions
Aureka ThemeScape	⊕				⊕	⊕	⊕		
ClearResearch	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
OmniViz	⊕			⊕	⊕	⊕	⊕		
Vivisimo					⊕				
Delphion Text Clustering					⊕				
VantagePoint	⊕	⊕	⊕	⊕	⊕	⊕			
Technology Watch		⊕		⊕		⊕			
M-CAM								⊕	
Aureka Citation Trees								⊕	
Delphion Citation Link								⊕	
Metrics Group Citation Bridge								⊕	
Wisdomain Citation Module								⊕	
Knowledgist		⊕							⊕
MS Excel	⊕	⊕							
Aureka Reporting		⊕	⊕						
Wisdomain Analysis Module		⊕	⊕						
Delphion PatLab II		⊕	⊕						
SciFinder		⊕							
SciFinder Panorama			⊕						

EXAMPLE

- LecturePatent_22\Lecture Patinformatics\NHR-RKS-JIPR153197-205.pdf

PATENT CITATION ANALYSIS

- Citation analysis is the important technique of patent analysis.
- Patent citation analysis is **a recent development which uses bibliometric techniques to analyse the wealth of patent citation information**.
- When a patent applicant cites “prior art” in the text of the patent, such citations are calledapplicant citations. Prior art is any evidence that your invention is already known.
- When the examiner grants a US patent, he will cite the key prior art on the front page of the granted patent; these citations are called examiner citations.
- The relationship between the patents that cite one another are shown in the form of hyperbolic trees
- This tool is very useful as far as analytics of patents are concerned, by using this inbuilt tool a user/searcher can visualize and carry out the patent mapping of backward and forward citations mentioned in the patent records. A user/searcher can map backward and forward references or both at the same time and make citation tree and generations

- Any patent cited six or more times would be in the the most highly cited 10% of all patents.
- The citation-based patent study is called patent citation analysis which links patents in the same way that science citation links the references in scientific papers
- The key idea behind patent citation analysis is that when a patent is very highly cited (i.e. cited in 5,
- 10, 20, or more subsequent patents) then that highly cited patent is likely to contain an important technological advance - an advance that many later patents are built upon.
- patent citation analysis has been used as a measure of technological quality and influence and in studying diffusion of technological information.
- Patent co-citation analysis has been used to “map” the inter-related development of technical fields
- Example: Engelsman and Van Raan used bibliometric (in particular patent-based) methods and techniques to develop a cartography of technology. Two types of maps are presented: co-word maps and co-classification maps. Both types of maps have been constructed for the entire domain of technology (the macro-level), i.e. the ensemble of all fields of technology in their mutual relations.
- LecturePatent_22\Lecture Patinformatics\E.C. Engelsman; A.F.J. van Raan (1994). A patent-based cartography of technology..pdf

PATENT CITATIONS

- Patent citations are the count of citations of a patent in subsequent patent or non-patent literature.
- Patent citations allow the analyst to assess the quality and impact of cited material, as well as the linkages between cited and citing countries, between cited and citing companies, and between cited and citing scientific and technological areas
- Sections include:
- ***Identifications off leading-edge technological activity: Example:*** In one study patents associated with outstanding new products have been shown to be highly cited. automotive, electronics, photographic and pharmaceutical industries have demonstrated the areas of leading-edge Japanese technological activity based on the study of the most highly cited US patents of Japanese origin
- ***Measurement of national citation performance:*** CPR= % of a country's patents in the top decile / % of all patents in the top decile A ratio of > 1 is good, a ratio of < 1 is relatively 10
- ***Technological mapping:*** Maps" of technological domains are constructed by examining the inter-relations of heavily cited patents. This helps in assessing relative position of companies (or countries) within each domain.
- ***Competitive intelligence :*** Patent citation analysis is a useful competitive intelligence tool.
- ***Linkages to science:*** Patent citations may quantify the dependence of technology on science and provide an indicator of how near a set of patents is to science.
- Measure of foreign dependence Foreign patents cited may be used to provide linkage between citing patents and foreign inventions. In a validation study of patent citations as indicators of foreign dependence, a high degree of association was found between the peer ranking and bibliometric measures.

INDICATORS

- The technological indicators based on patent citation analysis reported in the literature
  - citations per patent: There are valid reasons for believing that highly cited patents are patents of more than average technological impact.
  - highly cited patents: Highly cited material, such as highly cited earlier US patents, tend to be important, seminal inventions
  - non-patent links: The “other publications” cited in a patent may be used to quantify the dependence of technology on science. The number of these other publications or “other references” are its non-patent links.
  - technical impact index This is the percent of patents in a particular period which are in the most highly cited 10% of all patents. The expected value of the TII has been normalised to equal 1
  - current impact index: The current impact index is a measure of how often patents are cited in other patents, which shows how frequently they are used as the foundation for other inventions. It is calculated on the basis of how highly cited the past five years’ patents in a particular field are.
  - technology cycle time: This is the median age, in years, of the earlier US patents reference cited in the company’s new patents.
  - technology strength: The technology strength is a product of the number of company patents and current impact index. That is, the number of patents multiplied by the current impact index.

- **Why Analyze Patent Information?**
- Patents are intellectual property rights for the protection of an invention in the territories of individual jurisdictions which may be granted in exchange for disclosure of the invention
- **Types of Patent Documents and Publication Policies**
- The specific rules for applying for a patent and for processing patent applications, including their publishing, can be different and should be considered on a jurisdiction to jurisdiction basis.
- Patentable subject matter is also different between various jurisdictions. Most jurisdictions have a system in place with substantive examination, i.e. where the claimed technical subject matter is examined whether it meets certain conditions for patentability, such as novelty, inventive step and industrial applicability.
- **Pre-grant Publications**
- The process of generating a patent right starts with the first filing of an application with a national or regional patent office or with WIPO (namely, the International Bureau of the PCT). This office is sometimes referred to in patent analytics as the Office of First Filing (OFF).
- Some patent authorities keep patent applications secret until they are granted, but most authorities publish patent applications 18 months after their filing date, or the priority date, if the office is an OSF. These documents are called pre-grant applications and they don’t represent a granted right in their present form but may be granted in the future.

- Understanding the difference between granted patents and pre-grant applications is critical for interpreting their impact on a field. In the development of analytics associated with PLRs it is good practice to separate pre-grant applications from granted patents when conducting an analysis, e.g. by using kind codes (see section 4.2.4 below). The implications of pre-grant applications are significantly different than what can be implied from a granted patent and they should be considered separately, or at the very least, identified as a different type of document when visualizing a result.
- **PCT Applications**
- While pre-grant publications of applications are usually associated with specific jurisdictions and therefore may represent an indicator where innovation takes place or where patent protection is sought, there is also a special type of patent application that facilitates the filing of applications in many jurisdictions simultaneously.
- **Granted Patents**
- Publications of granted patents are of particular importance in comparison to publications of unexamined applications because the grant asserts that the invention disclosed in the application is indeed new and inventive over the known prior art. A grant can therefore be taken as a quality indicator for innovation activities.
- **Post-Grant Documents**
- There are a number of additional patent documents that may be published after the publication of the grant of a patent. The most important ones are publications following reexamination or oppositionprocedures that were initiated by third parties after the publication of the grant. If, as a result of these procedures, the scope of protection was restricted, a new publication would be made including the modified claims. That is the case, for instance, for the EPO, where if the patent is maintained in anamended form, a new patent specification is being published.

- **Kind Codes**
- It was already repeatedly mentioned that different publication stages of an application are usually distinguished by kind codes (e.g., A1, A2, A3, B1,..) which are part of the document publication numbers.
- **Components of Patent Documents**
- While patents documents contain a good deal of raw text, they are referred to as semi-structured, since they have a number of sections that are found in almost every document, regardless of its country of origin. At a high level these sections of a patent document are represented by a Front Page with bibliographic data, a Description (Disclosure) and a Claims section. Within each of these high-level sections there are subsections that provide specific information about the particular document. These subsections are typically segmented into individual fields when the documents are processed for electronic delivery or the generation of databases.
- **Front Page and Bibliographic Data (Metadata)**
- The term “bibliographic data” denotes the various data normally appearing on the first page of a patent or industrial design document or in a comprehensive entry in an official gazette concerning granted patents, industrial design or trademark registrations or the corresponding applications. Such data comprise document identification data, data on the domestic filing of the application, priority data, publication data, classification data and other concise data relating to the technical content of the document or of the entry in the official gazette.
- The majority of the statistical analysis conducted on patent collections takes place using data collected from the bibliographic fields within them. Many of these fields contain categorized text or numbers and thus are readily applicable to statistical analysis

- **Applicant/Assignee**
- *The applicant is the entity or person which or who presents ("files") an application for the grant of an industrial property right (e.g., a patent application, or an application for the registration of a trademark) in an industrial property office, or in whose name an agent (representative) files such an application*
- In general, the applicant is the inventor, but it may also be the employee or the person to who the inventor assigned his/her right to the invention (assignee). Ordinarily, this will be a company or organization, but can be the inventors when the rights associated with the invention are not transferred, or assigned, to a different entity.
- **Dates**
- Dates correspond to the timing of significant events in the lifecycle of a patent application. The three most significant patent related dates are the priority, the filing and the publication dates.
- The filing or application date is determined by the patent authority that receives the application if certain minimum requirements are fulfilled, which actually differ from jurisdiction to jurisdiction. The filing date may therefore differ from the date the applicant lodges the application with the patenting authority.
- **Priority Data**
- The part of the bibliographic data (normally published on the first page of a patent document) identifying the earlier patent application(s) on the basis of which a so-called priority right has been claimed
- **Classifications**
- In patent information and documentation matters "classification" means a specific system which subdivides technology into distinct units. A classification symbol is defined for each of those units. The classification symbol designating the unit into which the invention falls is usually printed on the first page of the corresponding patent document and recorded in databases as part of the bibliographic data.

- **Citations**
- During the prosecution of a patent application, an examiner will look for prior-art related to the novelty, obviousness, or an inventive step, associated with an invention. When references of this nature are discovered they are cited within the document during different publication stages. Usually within a search report that accompanies the document.
- **Description (Disclosure)**
- The description of the invention is one of the essential parts of certain kinds of patent documents, e.g., patent applications or patents. It usually specifies the technical field to which the invention relates, includes a brief summary of the technical background of the invention and describes the essential features of the invention with reference to any accompanying drawings.
- The patent system is built on the principle that protection of an invention is granted in exchange for disclosure of the invention in order to spur further innovation. The disclosure of the invention has to be clear and sufficient enough to enable experts in the field to carry out the invention.
- **Claims**
- The part of a patent document which defines the matter for which protection is sought or granted
- Each patent application has to include at least one claim. The first claim, the so-called main claim, is supposed to include all technical features of the invention that are essential to solve the technical problem that led to the invention and which is supposed to be solved by the invention.

- **Publicly Accessible Supplementary Information Associated with Patent Applications**
- Besides looking at the structure of individual patent documents, it is also important to understand that patents exist within an infrastructure of additional information associated with the development of the document, what occurs with it as it matures, and how it relates to other documents that are associated with it.
- Most patent issuing authorities keep all of this information in one place, but the United States provides
- three different databases for finding this data:
- • Public PAIR – US Case Histories – http://portal.uspto.gov/pair/PublicPair
- • USPTO Assignments – Re-assignments - http://assignments.uspto.gov/assignments/q?db=pat
- • US Patent Maintenance Fees – https://ramps.uspto.gov/eram/patentMaintFees.do
- **File Wrappers and Prosecution History**
- Patents, in the process of being examined, are prosecuted at a patent issuing authority. During the process, Office Actions and other procedural items take place between the patent office, or examiner, and the applicant with their attorneys.
- An example of understanding patent file histories in the United States can be found at
- http://www.tms.org/pubs/journals/JOM/matters/matters-0302.html.

- **Maintenance Information**
- Maintenance fees or renewal fees are fees that are paid to maintain a granted patent in-force. Some patent laws require the payment of maintenance fees for pending patent applications. Not all patent laws require the payment of maintenance fees and different laws provide different regulations concerning not only the amount payable but also the regularity of the payments. In countries where maintenance fees are to be paid annually, they are sometimes called patent annuities
- **Assignment Information**
- An assignment involves the sale and transfer of ownership of a patent by the assignor to the assignee. The assignee is the entity that is the recipient of a transfer of a patent application, patent, trademark application or trademark registration from its owner on record (assignor).
- **Patent Infringement and Litigation**
- Patents, by definition, are a right to exclude others from making, using, offering for sale, or selling an invention in the jurisdiction covered by an in-force document. After a patent has been granted and when the patent owner believes that an organization is performing one of these acts, with an invention covered by one of their patents, they can initiate litigation in the form of a patent infringement lawsuit
- **Patent Families**
- Due to the territorial character of the patent system worldwide, patents protection is sought in individual jurisdictions. The Paris Convention of 1883 facilitates the filing in different jurisdictions by claiming priority rights derived from earlier filings (at the offices of first filing - OFF). These priority claims lead to relations between different national patent applications, so called patent family relations. Since the Paris convention expressly permits the claiming of more than one priority rather complex family relations may exist depending on whether two applications share priorities in full,partially or only indirectly, i.e. through other ones.

- detailed example of the different definitions of patent families can be found at the URL below:
- http://www.epo.org/searching/essentials/patent-families/definitions.html
- In addition to the general concepts of simple and extended patent families, various database producers have created their own definitions of patent families for organizing patent documents within their collections, such as FamPat. The Intellogist wiki provides definitions for the major providers as well as additional general definitions for patent families besides simple and extended.
- http://www.intellogist.com/wiki/Patent_Families
- **Sources of Patent Information**
- The decision on what patent information source will be used is an important one for an analyst to consider as they are initiating a project. The cost of acquiring data to analyze should be weighed against the time it will take to work with the data, and its comprehensiveness
- While only the patenting authorities themselves generate **authoritative** patent data (**primary** sources), there are a number of different **secondary** sources for patent documents and information that are usually used to generate PLRs because they include patent information of more than just one jurisdiction which they have obtained from different primary sources.

- **Primary Sources: Patent Authorities**
- Each patent jurisdiction defines its publication policies and the authority in charge of producing the official patent related publications and providing access to other information like legal status data or the public part of the file wrapper
- Many patent authorities around the world have websites with data services that allow the general public to search and retrieve the respective patent documents
- Some of the features associated with these primary sites include:
- These collections are generally available to search for no cost; very few jurisdictions permit access to full publications only for a fee. Basic bibliographic data are generally accessible for free.
- Some of the offices (e.g. the United States Patent Office) separate the searching of applications from granted patents but most allow the user to search both simultaneously
- Some of the authorities maintain separate patent register databases that provide information on the most recent legal status of pending applications or granted patents, or permit file inspection, in addition to the official document publication services.
- Many of the primary sources allow searching in an English interface
- Search syntax and functionality varies from site to site so individual search strategies need to be developed for each such source.
- allow for bulk downloading of patent documents discovered during a search, while others only allow small numbers or single documents to be downloaded
- most sites do not allow individual patent data fields to be exported (structured data) or, if they do, the number of fields available is limited.

- **Free Secondary Sources**
- A few patent authorities maintain secondary patent databases which allow searching for patents from several countries together. These are mainly PATENTSCOPE from WIPO, Esp@cenet from the European Patent Office, or DEPATISNET from the German Patent and Trademark Office. Google, The Patent Lens, and the European Patent Office.
- Rudimentary analytics tools can be found on a few of the sites but this functionality is normally left to the commercial sources. For instance, TheLens from Cambia, and PatentInspiration from CREAX, and PATENTSCOPE offer some statistical analysis and visualization features.
- **Commercial Sources**
- Commercial sources of patent information have been available for over a century. What started as abstracting and indexing services covering patents from a handful of countries and, on a small variety of topics, has developed into a large business with many significant players Clarivate Analytics, CPA Global, and Lexis Nexis.

- They offer **Enhanced content, A “one-stop-shop” for searching, analysis and dissemination** – several of the major commercial providers allow an analyst to search, refine, review, analyze and share collections and output within the same system. **Flexibility in exporting data** – Commercial sources generally have a higher limit on the number of records available for export. **Additional tools for refining data collections** – as will be discussed in subsequent sections of the guidelines, patent data can contain errors, such as typos in patent assignee names, or redundancies, such as the same invention being represented in different countries
- **Reports Associated with Patent Information**
- Due to the critical nature of patent documents and the information associated with them, reports related to patent information are used in a variety of different business contexts. There are different reports affiliated with providing information on patent data in these different environments.

- **Patent Landscape**
- There is no single definition or common understanding for a patent landscape report. There are various approaches, some of which broader, covering even Freedom to Operate elements and other, non-patent related data, such as market analysis, while others much narrower, with certain understandings being that a patent landscape is identical to a patent map
- **Patent Map**
- While the name patent map sounds similar to a patent landscape, a patent map generally represents a graphical representation of a data collection that borrows characteristics of cartography. Maps are usually focused on a single attribute associated with a data collection such as the classification of documents based on the topics covered within them.
- **Watch or Alerts**
- A patent watch is a process for monitoring newly issued patents, as well as possibly pending patent applications, to assess whether any of these patent documents might be of interest
- **Freedom-to-Operate / Clearance**
- In this type of report, which involves an organization asking for a legal opinion on whether a product they are planning on shipping will infringe any existing patents before they launch. The search involved is very specific since it is country specific and usually only applies to in-force granted patents and their claims

- **Patentability / Prior-Art**
- This type of report is usually performed in the legal context of determining if a new invention is eligible for patent protection and determining how broadly the claims for the new invention can be written. This type of report can cover both patent and non-patent literature and is typically looking for references that were published before the filing date of the invention in question
- **Validity**
- Validity reports provide the results of the largest and most comprehensive of all patent searches.
- These reports are almost always associated with large sums of money and critical business decisions and as such need to be as comprehensive as possible. This report shares similar characteristics to Patentability but is normally far more comprehensive since there is typically much more at stake when this sort of report is asked for.
- The object of the search involved with this report is to identify prior art references, which will allow a granted patent to be made invalid or revoked during a particular proceeding before the particular patent office of interest or during a court proceeding

PATENT LANDSCAPE REPORTS

- Patent Landscape Reports (PLRs) support informed decision-making, and are designed to efficiently address the concerns associated with making high stakes decisions in various areas of technology, increasing the related degree of confidence
- With the institution of patent analytics, and PLRs, it is possible for these critical decisions to be made with data-driven, evidence-based approaches that deliver informed choices, and mitigate the associated to the decision risks.
- PLRs can be used as instruments to inform public policy makers in strategic decisions to related to R&D investment, prioritization, technology transfer or local manufacturing.
- It is critically important to make certain that a PLR is prepared properly in order to ensure that the insight it provides is accurate, and directed towards the key issues associated with technological implementation.

OBJECTIVES AND MOTIVATIONS FOR GENERATING PATENT LANDSCAPE REPORTS

- Producing a Patent Landscape Report (PLR) can be a time intensive and expensive process. Devoting the resources necessary to generate a PLR is often tied to a business objective, e.g., where an organization is preparing to make a significant monetary or headcount investment in developing or moving into a technology area.
- The issues associated with public policy decisions, initiated by government agencies, are usually different from the decisions that are important to corporate entities, and their stakeholders. The analyses of patent information, and the generation of PLRs, are increasingly required by both types of organizations, in order to understand a technological area. Understanding how the decisions differ between these two types of entities allows the analyst to tailor their report in order to most efficiently meet the needs of the respective audiences.

- **Global Efforts**
- *Good quality information about patenting activity is an essential input for some of the most critical international policy debates today. Yet patent information is unavoidably complex, constantly evolving, and difficult to capture in readily accessible form for a non-specialized audience*
- Facilitate access to user-friendly global databases which contain public information on the administrative status of health-related patents. This includes supporting existing efforts for determining the patent status of health products, and to Promote further development of such global databases including, if necessary, compiling, maintaining and updating such global databases
- **Technology transfer and licensing**
- In order to assist industries residing in their countries, several national governments have recently started purchasing patent assets on behalf of the organizations that manufacture products in their jurisdictions.
- *Korea's Intellectual Discovery, which was started in 2010 amid government fears that domestic companies were losing key patents that could be used against them by foreign companies, has a $140 million government commitment.*
- PLRs are often used by organizations exploring technology transfer, and licensing in order to understand what other organizations have invested in a particular area. If another organization has invested in a technology, especially if the investment was made a few years earlier, there is a higher likelihood that they will be receptive to hearing about new developments, and potentially acquiring or licensing the technology.

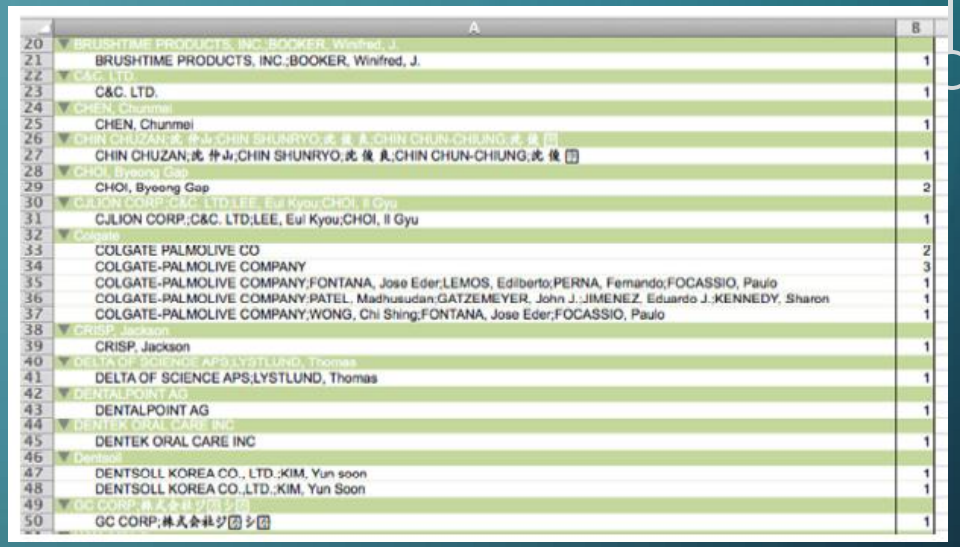
- **Research and development (R&D) decision-making**
- the use of PLRs can influence decisions around investments in academic and non-profit funding for the creation of economically favorable technologies, which can increase the Gross Domestic Product of a country.
- **Business or corporate uses**
- While there are some overlaps between the uses of patent analytics, and PLRs for governmental policy decision-making, the circumstances under which these tools are used for business, or corporate situations can be somewhat different
- **Competitor monitoring**
- Competition is an inherently business related concept and it is nearly impossible to find a successful industry in which there is not some form of competition between organizations looking to gain an advantage over other companies in a space.
- **Technology monitoring**
- Corporate entities usually associate themselves with specific technological areas.
- **Mergers and acquisitions**
- Technology savvy organizations will look at innovative benefit and fit, in addition to the more traditional economic and market driven factors, when deciding on whether to acquire, or merge with another company. A technology based due diligence assessment is necessary to ensure that tw groups will be compatible with one another and is often used to determine how much technological overlap there is between them.

- **Who is the report intended for?**
- In order for a PLR to be impactful it has to be read by a decision maker in the first place. The two primary considerations in determining whether a PLR will be read, and better yet acted upon, are based on the position your client holds within the organization, and their personal information intake profile.
- **How does it save the client time?**
- The most valuable commodity a PLR can provide a decision-maker is to save them time. PLRs take large amounts of raw, unprocessed data and generate insight, and intelligence based on the analysis conducted and, perhaps more importantly, the interpretation of the results by a skilled analyst
- **How will the user evaluate the effectiveness of the report?**
- As stated at the beginning of the chapter, the generation of a PLR is a time intensive and expensive endeavor, and it is embarked upon since an organization is preparing to make a significant monetary or headcount investment. Understanding how the user will evaluate the PLR is a critical component in ensuring that the end product will be satisfactory to them.

- **Data Cleanup and Grouping**
- Data cleanup and grouping are processes for the manual, or automatic standardization of terms or items, within a data field, to correct errors or inconsistencies, or to group synonymous entries.
- It is required by patent analysts in order to produce statistically relevant results. It is necessary since raw patent data is notoriously "messy" and requires cleanup or standardization to produce accurate results.
- Using a simple data cleanup example, 3M is listed in most patent assignee fields in a number of different ways including: 3M, 3M Inc., 3M Inc, and by its full name, Minnesota Mining and Manufacturing. All of these represent the single entity, and have to be grouped together, or standardized, in order to perform accurate statistics reflecting the total impact of the organization.
- Grouping these terms together into a single entry is the essence of the data cleanup task.
- **Patent Assignee Cleanup** – Perhaps the most often used cleanup task due to misspellings and alternative representations of company names.
- **Inventor Name Cleanup** – Misspellings are also common in this field, but just as frequently there are issues with whether an individual uses their full first name, their middle name, or initial, and in the case of the last name, whether they have changed it due to marriage. Asian names, and the order in which they appear also cause problems when cleaning this field.

- **Technology Categories** – synonymous terms from text collections within patent data fields, such as the abstract, claims or examples, are grouped together to represent a single concept. Cancer, for instance, can be described using additional terms such as neoplasm. Grouping terms when creating technology categories are required to ensure that all relevant occurrences of a concept are captured during analysis.
- **Up Posting Classifications** – Technical subject matter is often classified by hierarchical lists of concepts, but analysts may not always want to use them at their most granular, or detailed level.
- When this is the case, more detailed classification can be grouped together, and collected as part of a higher-level entry. Using IPC classifications as an example, an analyst may group several subgroups together, provide them with a name that represents a meaningful concept, and then processes them as a single entity

- **Manual Methods**
- These methods require the analysts to work with each individual entry in a data field. It requires that the analyst knows the relationship between one entry and another, and to be able to decide, if they are in fact, the same entry.
- **Pivot Tables –** Pivot Table functionality allows a user to select one or more rows in a spreadsheet and group them together.
- **Drag and Drop or Manual Grouping –** Some tools provide a list of entries within a field and allow the user to either drop one entry on to the top of another to create a grouping, or allow multiple entries to be selected, and then grouped by right-clicking or pushing a button.



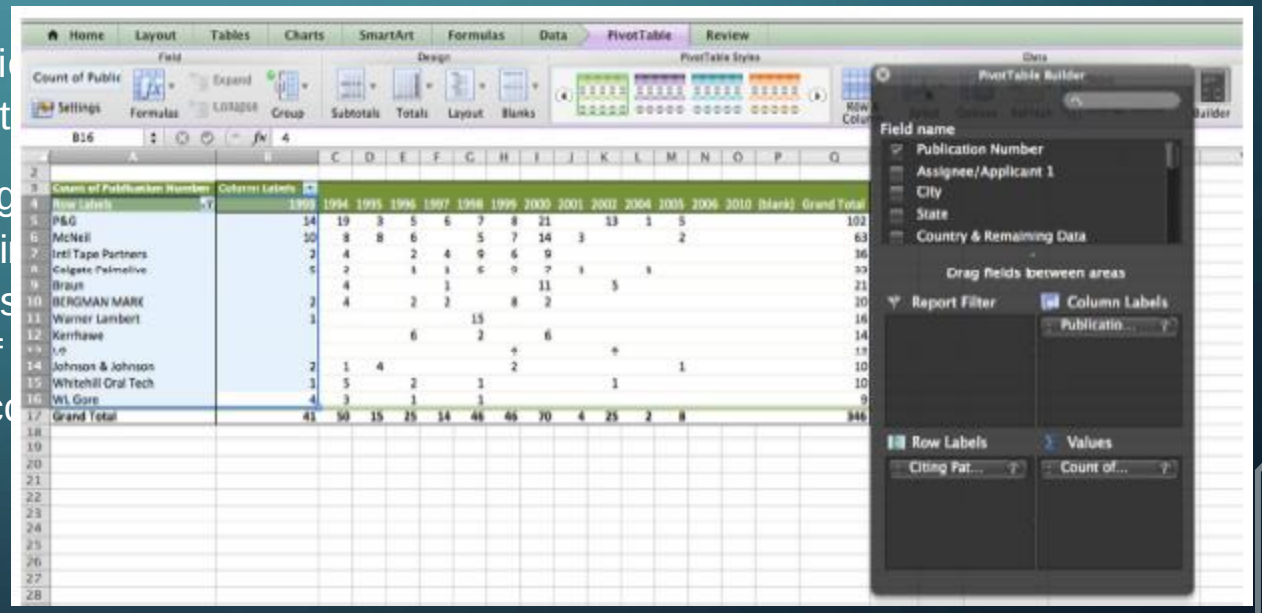
	A	B
20	BRUSHTIME PRODUCTS, INC.;BOOKER, Winifred, J.	
21	BRUSHTIME PRODUCTS, INC.;BOOKER, Winifred, J.	1
22	C&C. LTD.	
23	C&C. LTD.	1
24	CHEN, Chunmei	
25	CHEN, Chunmei	1
26	CHIN CHUZAN;沈 仲Jr;CHIN SHUNRYO;沈 俊 良;CHIN CHUN-CHIUNG;沈 俊	
27	CHIN CHUZAN;沈 仲Jr;CHIN SHUNRYO;沈 俊 良;CHIN CHUN-CHIUNG;沈 俊	1
28	CHOI, Byeong Gap	
29	CHOI, Byeong Gap	2
30	CILION 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	
50	GC CORP;株式会社ジ	1

- **Automated Methods**
- These methods allow the processing of an entire field of data based on the use of an algorithm, or an agreed upon collection of facts. While not as time consuming as manual methods the accuracy of the method is only as good as the algorithm used, or the knowledge of the individuals who built the collection of facts.
- **Fuzzy Logic –** An algorithm looks at the string of characters associated with an entry and determines the likelihood that two items represent the same entry based on how similar they are to one another
- **Bootstrapping –** This method involves the collection of lists of standardized items that can be used to group together entries consistently. The organizations that build databases will often use these lists to make certain that they consistently apply the same name to an entry regardless of how it may have been used in the source material. Lookup tables can be used to transform raw input into standardized entries with large collections.

- STN AnaVist – provides manual grouping which can be saved by user
- VantagePoint – allows for the creation of user defined thesaurus and fuzzy logic grouping
- Orbit.com – provides manual grouping which can be saved by user, online
- Microsoft Excel – includes Pivot Table functionality for manual grouping

- **List Generation**
- List generation is a statistical method that provides counts of various patent related metrics within individual data fields. Identifying the top ten inventors, for instance, is an example of generating a list of the most frequently occurring inventors in a patent collection.
- Once the list is generated, the rows and columns representing the list can be selected, and a bar chart column chart visualizing the data contained in the list
- **Co-Occurrence Matrices**
- Counting or comparing entries within a single data field is accomplished by generating a list, as covered in the previous section.
- Frequently, an analyst will want to provide additi
incorporating an additional variable or field into t
- For instance, it is useful to understand which org
patent documents in a technical field, but more i
organizations filed the patent documents in ques
organizations will have a similar total number of
them much earlier than the other. Using a co-occ
context can be readily represented.

A	B
Company	Patents
P&G	11
Colgate	8
Kim Hye Kyung	5
Bosch Cerda	2
CHOI, Byeong Gap	2
Dentsoll	2
LEE, KYUE HYOO	2
LEE, KYUNG HO	2
MCNEIL PPC INC	2
MURPHY, PAUL	2
Profimed	2
Ranir	2
TIPHONNET JOEL	2



- **Clustering and Classification**.
- Clustering is normally associated with unsupervised methods of organizing document co… similarity comparison between documents. With a fixed number of clusters identified at t… collections that meet a threshold similarity component are grouped together.
- When it comes to clustering, the two most often used algorithms
- **K-means** – a method of cluster analysis, which aims to partition n observations into k clu… observation belongs to the cluster with the nearest mean.
- **Force Directed Placement** – At the most basic level the algorithm tries to place similarc… and dissimilar objects far apart. The process is achieved by moving the objects randomly… space via a technique similar to ‘simulated annealing.

Methods for classification

- **Artificial Neural Networks –** In computer science and related fieldsThey are usually pre… interconnected “neurons” that can compute values from inputs by feeding information thr…
- **Support Vector Machines** – supervised learning models with associated learning algor… and recognize patterns, used for classification and regression analysis. The basic SVM t… and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier

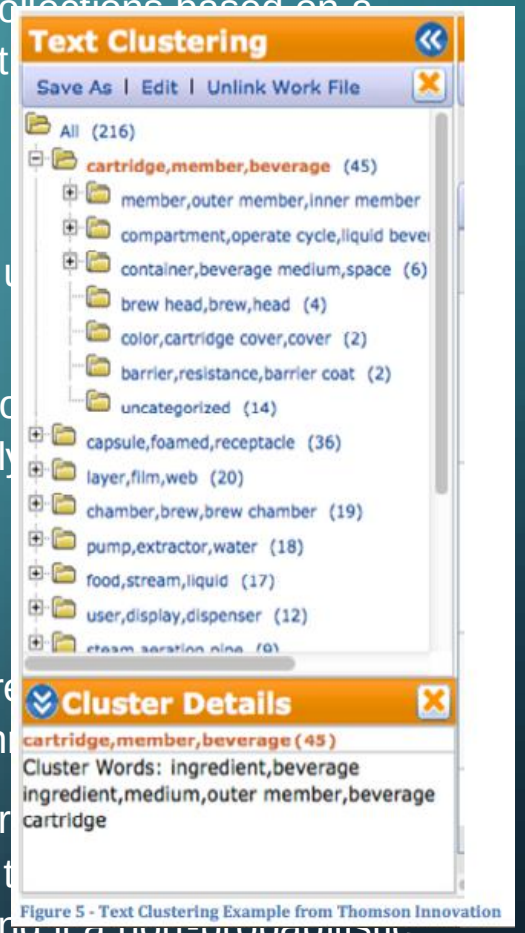
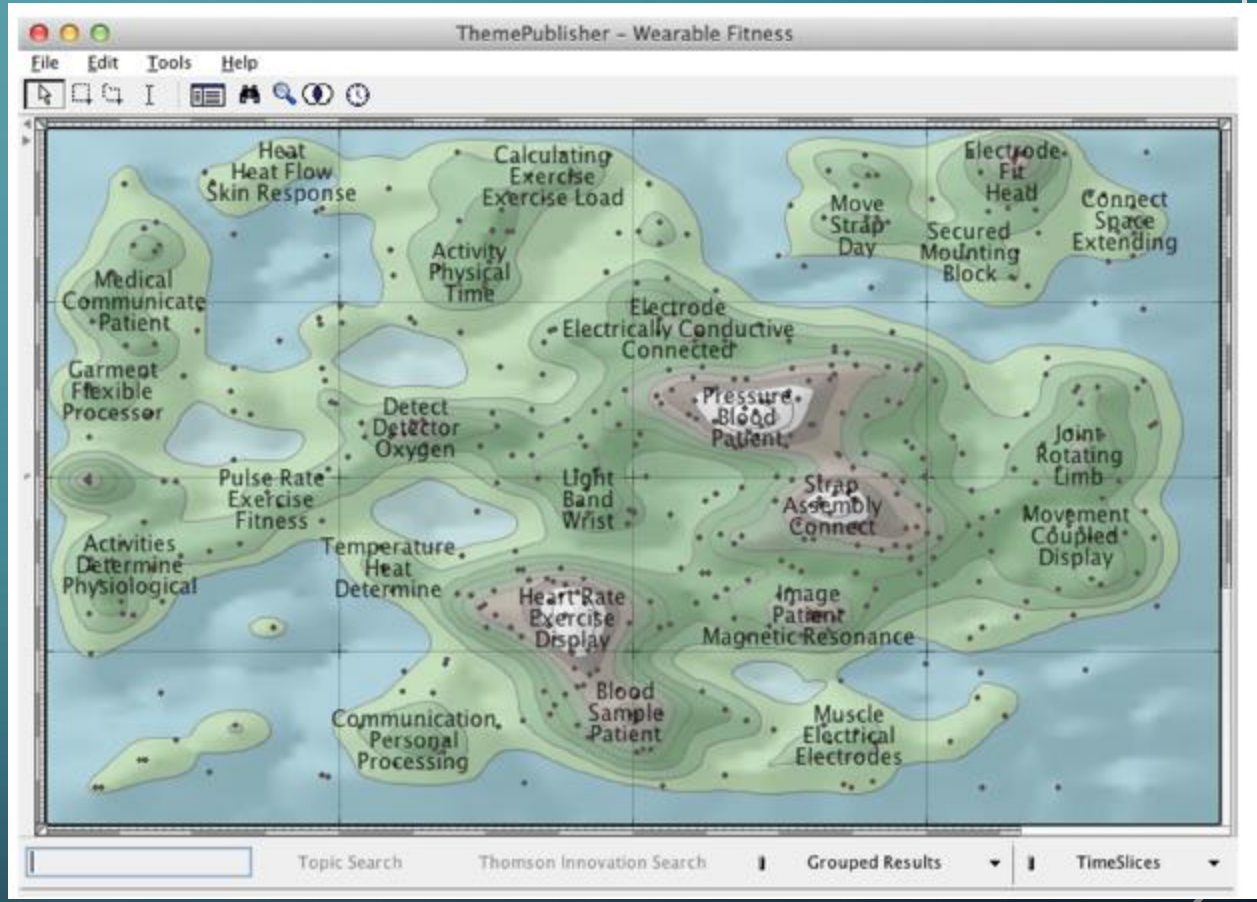


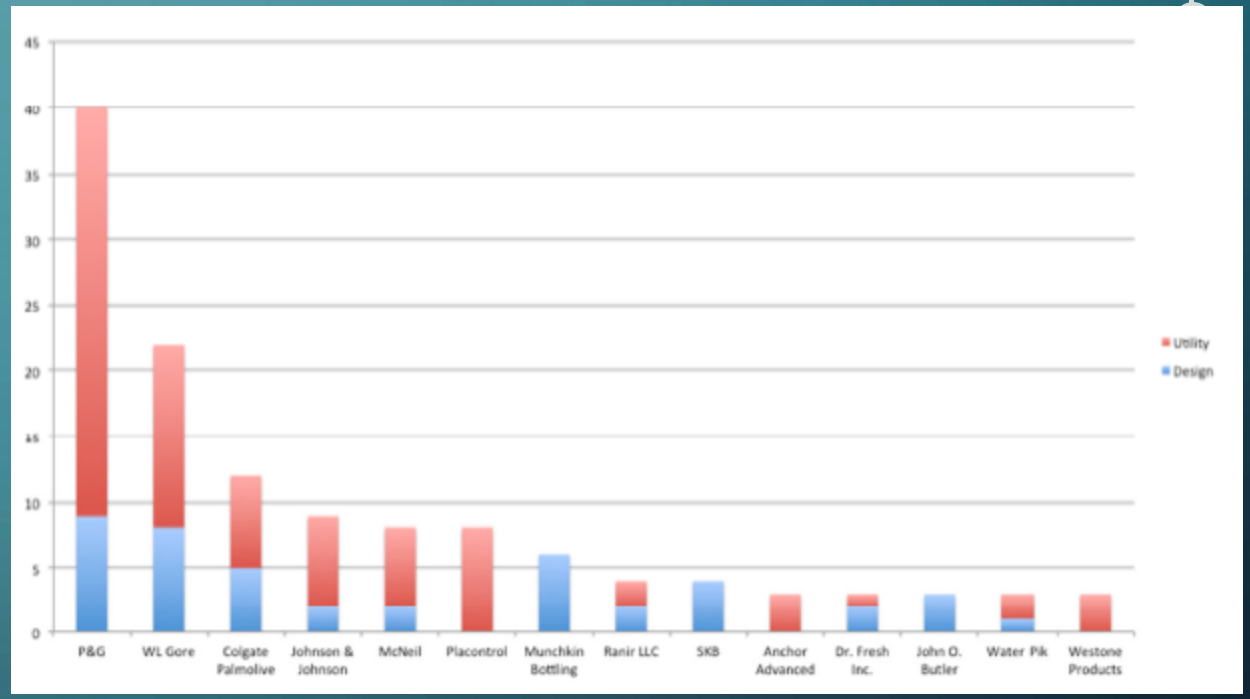
Figure 5 - Text Clustering Example from Thomson Innovation

- As applied to PLRs, and patent analytics, the most frequently used sources of text for both clustering and classification exercises come from patent classification codes, or from raw, or standardized text coming from the source document:
- **Classification Codes** – Intellectually assigned classification systems produce standardized codes that can be used as a means of categorizing documents that share similar subject matter.
- **Raw Text** – processed to identify concepts and phrases contained within specific sections of the source document, such as the abstract or claims. As with the clustering of structured data concepts, instead of codes, are used to group documents that share a high degree of overlap.
- **Indexing Terms** – producers of “abstract and indexing” databases normally produce hierarchical lists of indexing terms that are used to classify documents based on standardizedterms and phrases. Since these lists are standardized and intellectually assigned, they can be used for clustering exercises.

- Spatial Concept Mapping
- Mapping is related to clustering or classification exercises, where the systems involved take the document clusters or classes and arrange them in 2-dimensional space by considering the similarity of the documents relative to one another over the entire collection
- Kohonen Self Organizing Maps – a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map.Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space



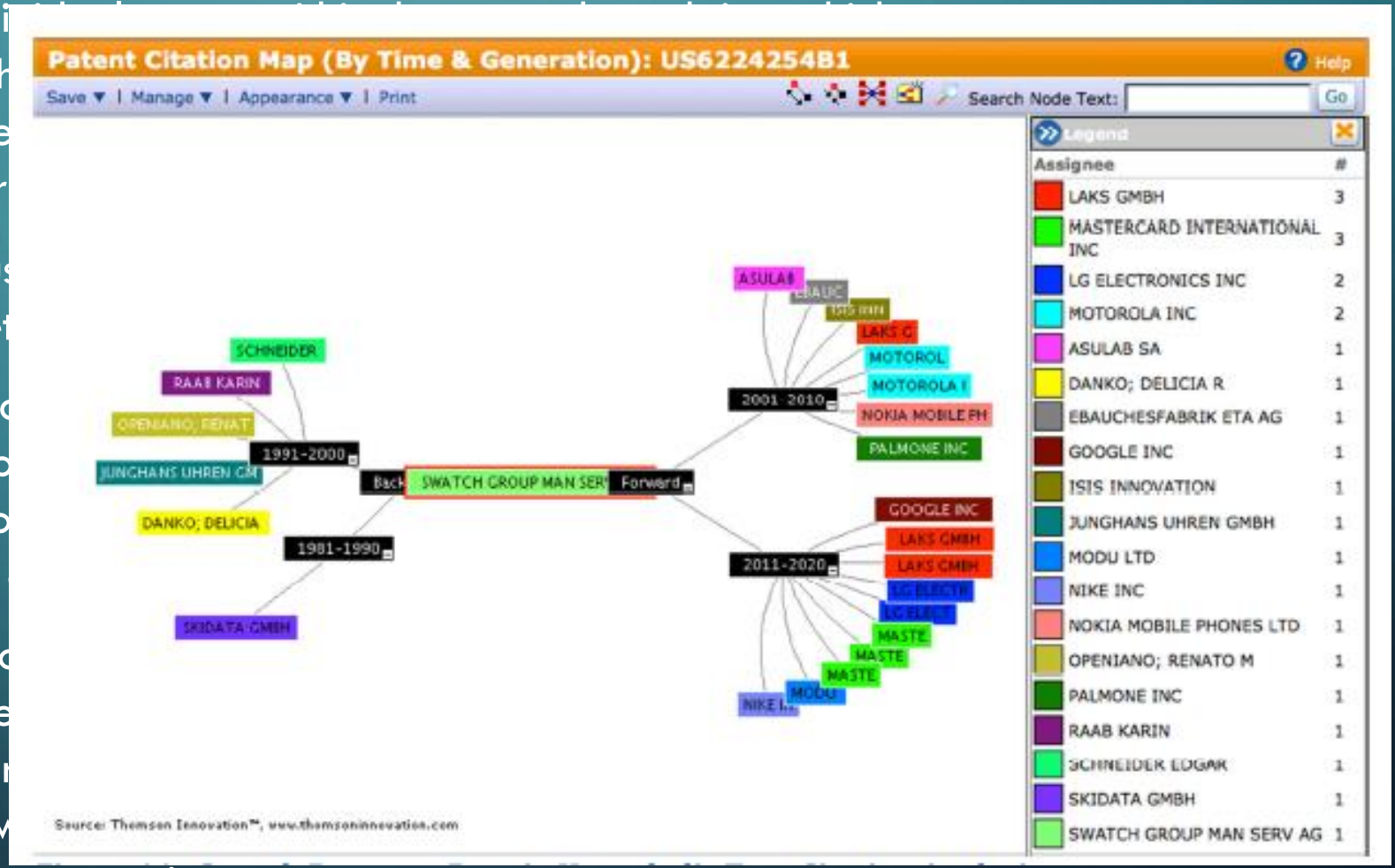
- **Layering or Stacking Information**
- Analyses looking at a single variable, or field, can be inefficient and lack context. Positioning two types of visualizations next to one another, or adding overlays to an analysis, allows the analyst to reference several attributes of a data set simultaneously without asking the client to refer back to previous illustrations.
- A stacked chart can be used to enhance a standard bar, or column chart when there is a need to explore a second variable with a small number of entries
- the mapping task was introduced, taking document clustering, and adding a similarity metric between documents to produce a graphic representation of how documents relate to one another based on shared concepts



- **Geographic Representation**
- *To georeference something means to define its existence in physical space. That is, establishing its location in terms of map projections or coordinate systems.*
- On most patent documents the physical addresses of both the applicants, and the inventors associate
  information
  placehold
  for the reg
- Recently,
  launched

Figure 10 - WO Applications from Inventors in Dublin, Ohio, USA Using World Patents, Mapped

- **Network Analysis**
- Network analysis is the viewing of relationships in terms of network theory70, consisting of nodes, representing indi
  relationships between th
  These networks are ofte
  points and ties are repr
- The two most common us
  inventor and citation net
- With respect to the conc
  examined, relevant prio
  documents. These citatio
  visualized in a network
- When an entire collectio
  network analysis can be
  of time. A network appr
  linear relationships betv
  relationships that skip a generation.



- While citation links exist between discrete
  the patents from an assignee together, a
  relationships that organizations have to c
  to provide a network graph
- Thomson Innovation – provides hyperboli
- Orbit.com – provides a network graph o
  module
- Intellixir – provides network analysis for
- AmberScope – generates citation networ

Figure 12 - Network Graph of Assignee Citation

- **Semantic Analysis**
- While there are several semantic techniques, such as latent semantic analysis, that are useful in the context of information retrieval, and patent searching, the primary form of analyzing patent documents using semantic information is the study of subject, action, object (SAO) triplets. SAO triplets utilize parts of language, such as nouns and verbs that are used to describe the teachings that the applicant wants to share. Key SAOs, within a patent document encapsulate the technical learnings contained in it.
- An example of this would be the treatment of asthma, where treatment is the action and asthma is the object. A potential solution to this problem would be the use of an allergy shot, where the allergy shot is the subject. This triplet would be extracted from a sentence like; an allergy shot was used to treat the asthm

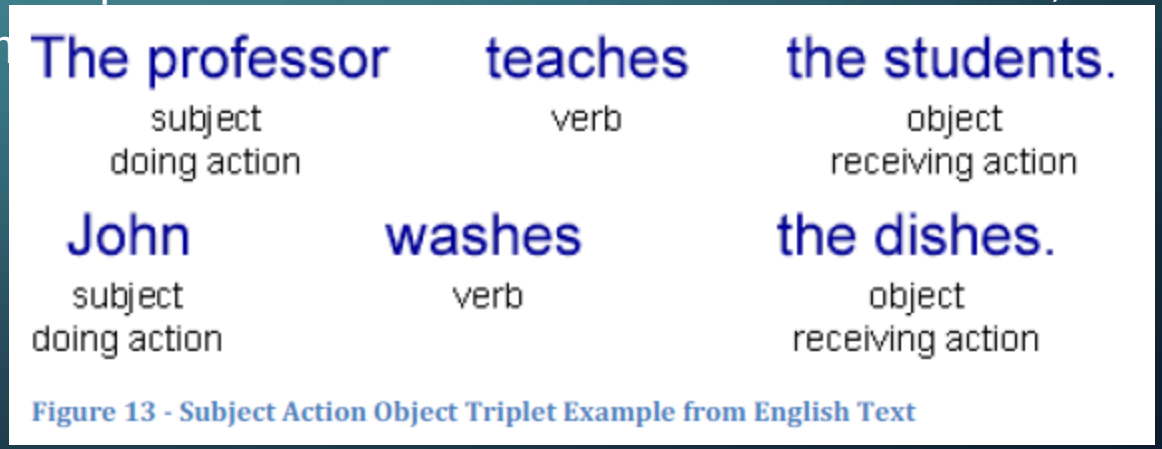


Figure 13 - Subject Action Object Triplet Example from English Text

- **Inclusion of Non-Patent Literature (NPL)**
- The word patent is included in the name Patent Landscape so the expectation is that patents will be the sole source of data being explored in these reports. Sometimes, however, the inclusion of NPL is required in order to meet the objectives of the report. For instance, due to the 18-month delay in the publication of patent documents, in most jurisdictions, truly cutting-edge developments can only be discovered when looking at NPL.
- Again, include NPL analyses when they are required to achieve the objectives of a PLR, and they provided a richer view of a topic area, but keep them separate from the analytical work being done on the patent documents
- **Preparing a Terms of Reference (TOR)**
- A Terms of Reference (ToR) document is not a standard practice in the private sector before preparing a Patent Landscape Report. That is comprehensible; Terms of Reference may seem to be too time-consuming or unnecessary for a task that may appear to be clear and straight-forward. Nevertheless, as the experience from the WIPO Patent Landscape Reports project has shown, drafting Terms of Reference before initiating a Patent Landscaping exercise is highly recommendable.
- The drafting of the Terms of Reference for a PLR should ideally reflect and follow discussions between recipient and provider of the report. That allows the assessment of the feasibility of certain requirements related to delivery timelines, pricing and the content of the report before the initiation of the projects, as often the lack of technical knowledge on patent information from the side of the commissioning party paired with possible lack of understanding about the contractors limitations and options in his deliverables cause misunderstandings and false expectations

Patent Landscape Report on .....

Technical Evaluation Criteria

#	Main Criteria	Candidate A	Candidate B	Candidate C	Candidate D
1	The Bidder's general understanding of the project	3	4	4	4
2	Specific experience relevant to project	4	3	3	4
3	Qualification and expertise of the project team	3	3	3	5
4	Understanding of the project deliverables and proposed approach for the preparation of the PLR	4	2	2	5
5	Responsiveness and compliance with legal requirements	3	3	3	3

Technical Evaluation Criteria - Weighted scores

#	Main Criteria	Weight	Max Score	Candidate A	Candidate B	Candidate C	Candidate D
1	The Bidder's general understanding of the project	3	15	9	12	12	12
2	Specific experience relevant to project	3	15	12	9	9	12
3	Qualification and expertise of the project team	6	30	18	18	18	30
4	Understanding of the project deliverables and proposed approach for the preparation of the PLR	6	30	24	12	12	30
5	Responsiveness and compliance with legal requirements	2	10	6	6	6	6
		Total	100	69	57	57	90

Financial Component

	Candidate A	Candidate B	Candidate C	Candidate D
Total cost in CHF	50000	33000	46895	64000
Score	22.9	34.7	23.4	17.9

Overall Weighted Scores and Ranking

#	Overall Scores	Weight	Candidate A	Candidate B	Candidate C	Candidate D
1	Technical Evaluation - Weighted Score	65%	44.9	37.1	37.1	58.5
2	Commercial Evaluation - Weighted Score	35%	8.0	12.1	8.2	6.3
		Overall Score	52.9	49.2	45.2	64.8
		Overall Rank	6	0	0	2

NOTE: Scoring shall be as follows: 5-Excellent; 4-Good; 3-Average; 2-Below average; 1-Poor; 0-not acceptab

Disqualified in Technical Evaluation (below minimum acceptable threshold X/5
Qualified in Technical Evaluation

Financial Evaluation

Lowest offer score 100 points and the rest are calculated as follows Lowest Offer/The Offer in Evaluation

- **Performing the Search**
- The data, in the case of PLRs, is patent information, and it is collected by means of various types of searches conducted in a variety of different databases. 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 thebusiness 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.

- **Determining Which Year Field will be Used**
- When conducting analytics, they are typically done on a year-to-year basis. Dates are provided for a number of milestones in the on-going lifecycle of a patent application, but working with specific dates is normally too granular, so years are used instead. There are three primary year choices associated with patent data, the application or filing, the grant or publication and the priority filing.

THANK YOU