

Manchester International Summer School on Emerging Technologies

8 - 13 June 2014, University of Manchester, United Kingdom

Chasing Future from Tech Mining

Analyzing patent topical information for potential technological opportunities

Jing Ma

School of Management & Economics,

Beijing Institute of Technology

Content

❖ **Background**

TOA and tech mining

❖ **Current study**

Analyzing patent topical information to identify technology pathways and potential opportunities

❖ **Next step**

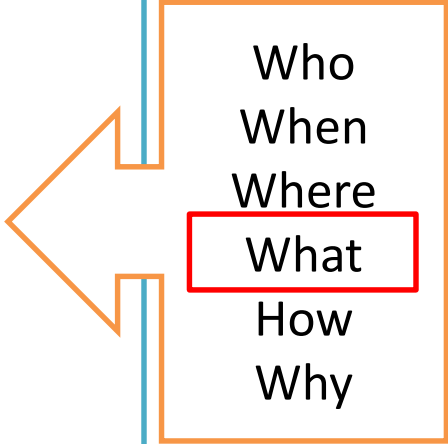
Literature-related discovery and innovation

Technology Opportunities Analysis (TOA) and Tech Mining

Exploiting literature information to see patterns, detecting associations, and opportunities.

Forecasting likely development paths for emerging technologies.

- Search and retrieve data (publications and patents).
- Profile the resulting search set.
- Extract latent relationships.
- Represent relationships graphically.
- Interpret the prospects.



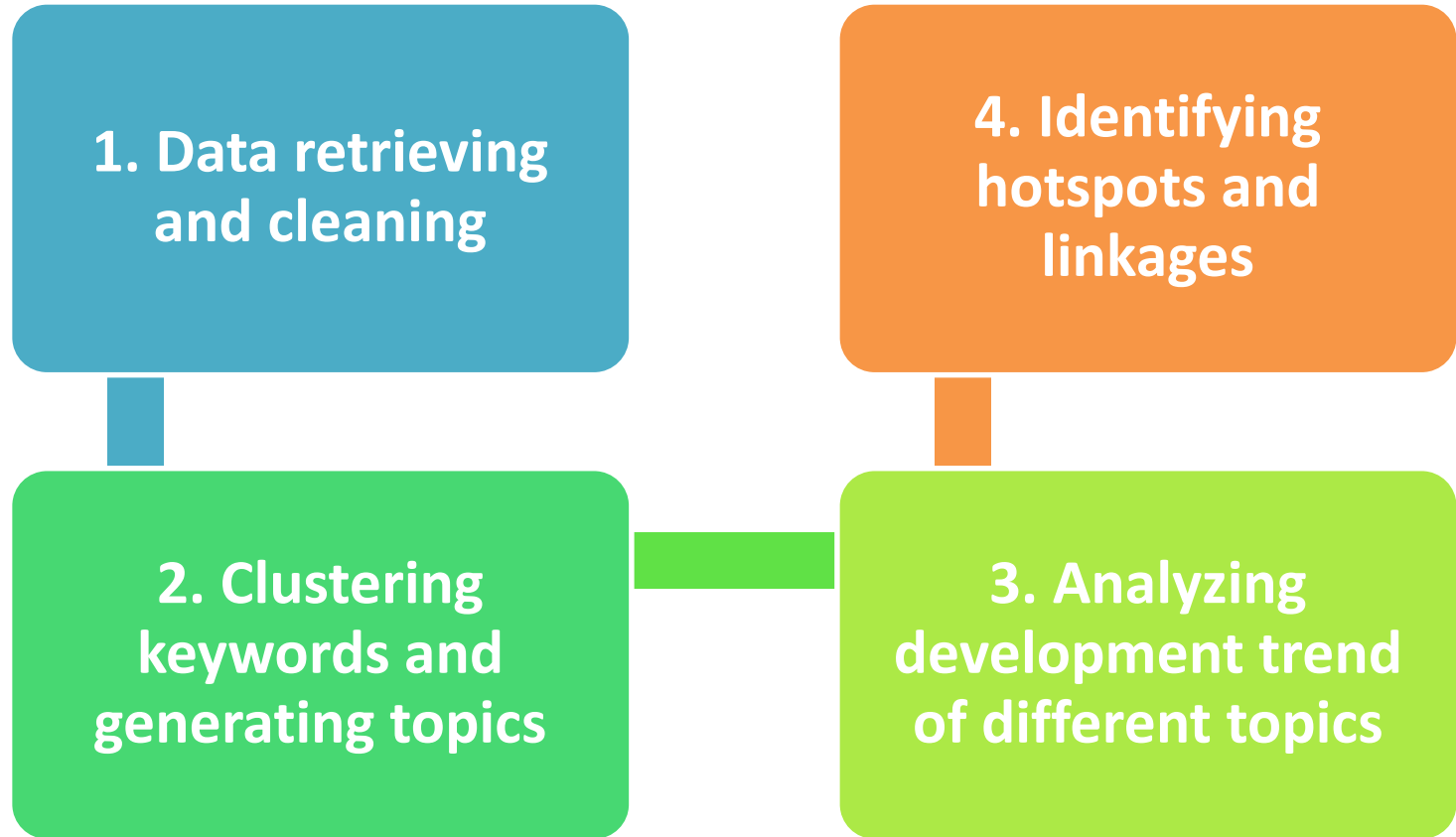
Who
When
Where
What
How
Why

Patent Topical Analysis

Comparing with publications, mining patent data is not overly effective.

- a) No keyword generated by authors, and the classification system is generic.
 - b) Descriptions in titles & abstracts are more obscure.
 - c) Novel phrases or words may be ignored due to their low frequencies.
- How to analyze patent text content and understand their topical information effectively?
 - How to identify potential innovation pathways and technology opportunities by using patent topical information?

Patent Topical Analysis



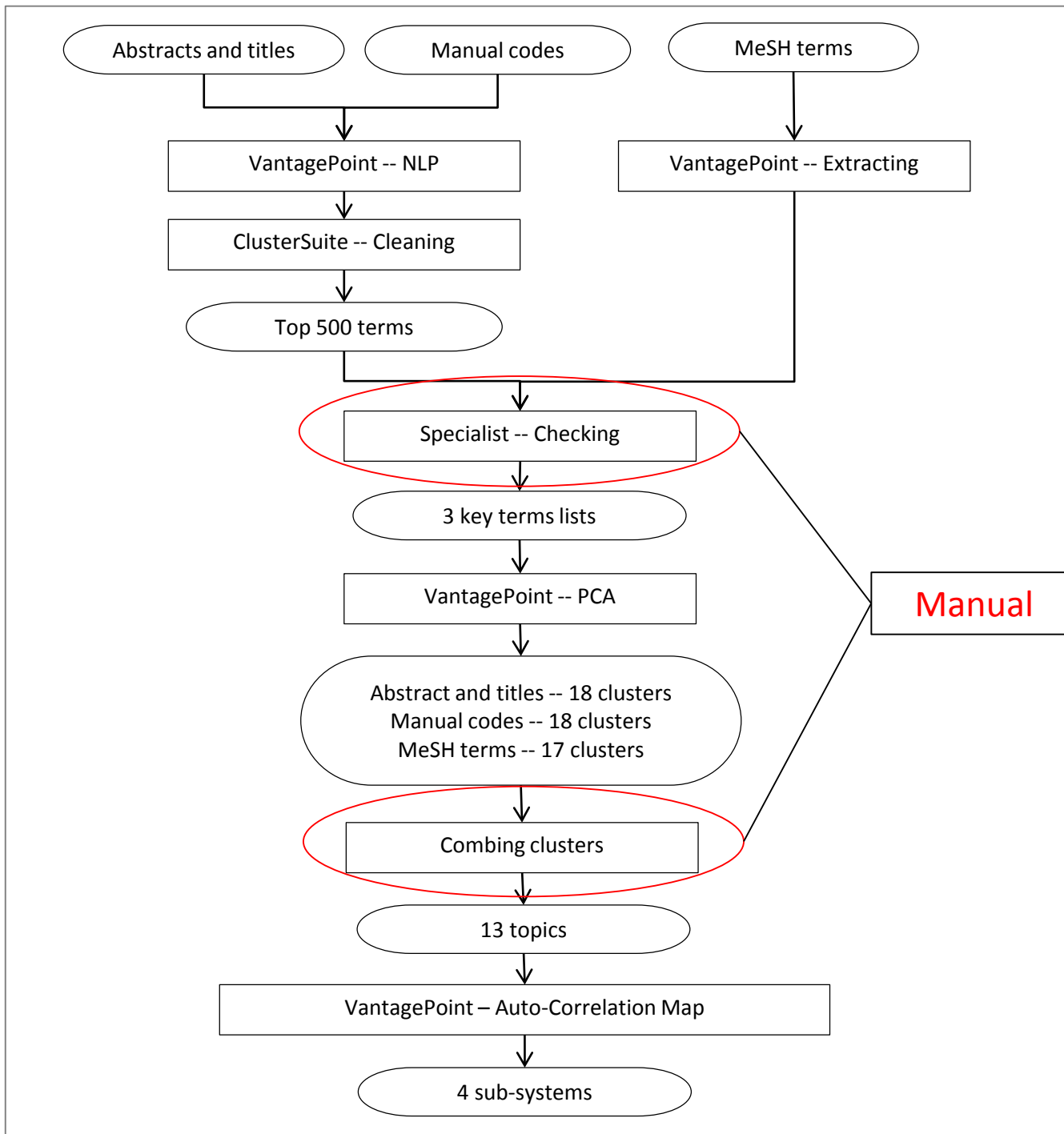
Current study – NEDD topical analysis

- Nano-Enabled Drug Delivery (NEDD).
- Derwent Innovation Index (DII) data
- From 2000-2012, 7906 records.
- Different ways to enhance topical intelligence from patent compilations.
 1. NLP noun phrases from abstracts and titles,
 2. patent classes (especially, Manual Codes), and
 3. external (imported) keyword list (Medical Subject Headings top 519 (MeSH) terms from MEDLINE (2012))

Patent example

Field Name	Value
Title	Lipid nanoparticle composition, for delivering e.g. proteins into cells, comprises short interfering nucleic acid molecules, dimethylamino-(cholesten-oxybutan-4-oxo)propane, cholesterol and polyethyleneglycol-diacylglycamide
Abstract	NOVELTY - Lipid nanoparticle composition (I) comprises one or more short interfering nucleic acid molecules, 3-dimethylamino-2-(cholest-5-en-3 beta -oxybutan-4-oxy)-1-(cis,cis-9,12-octadecadienoxy)propane, cholesterol and polyethyleneglycol-diacylglycamide. USE - (I) is useful in therapeutic, research, and diagnostic applications for delivering biologically active molecules (which is e.g. antibodies (e.g. monoclonal, chimeric and humanized), cholesterol, hormones, antivirals, peptides, proteins, chemotherapeutics, small molecules, lipids, vitamins, co-factors, nucleosides, polynucleotides, oligonucleotides, enzymatic nucleic acids, antisense nucleic acids, triplex forming oligonucleotides, 2,5-A chimeras, allozymes, aptamers, decoys and their analogs, and small nucleic acid molecules comprising short interfering nucleic acid, short interfering RNA, double-stranded RNA, micro-RNA, micro RNA inhibitors and short hairpin RNA molecules) into relevant cells, organs and/or tissues (e.g. such as skin, mucous membranes, vasculature tissues, gastrointestinal tissues, blood brain barrier tissues, ophthalmological tissues, pulmonary tissues, liver tissues, cardiac tissues or kidney tissues) in a subject or organism. (I) is useful for treating or preventing e.g.: a proliferative disease or cancer (e.g. acute myelogenous leukemia, meningiomas and Giant cell tumors); inflammatory disease (e.g. atherosclerosis, psoriasis and tuberculosis); transplant and/or tissue rejection (allograft rejection); autoimmune disease (e.g. multiple sclerosis, Crohn's disease and Addison's disease); age-related disease; infectious disease (e.g. virus, bacteria, fungus, prion and parasite); neurologic or neurodegenerative disease (e.g. Huntingtons disease, Angelman syndrome and Behcet's disease); metabolic disease (e.g. obesity, insulin resistance and diabetes); cardiovascular disease (e.g. coronary heart disease, aortic stenosis and myocardial infarction); respiratory disease (e.g. asthma, pulmonary vasoconstriction and respiratory distress syndrome); ocular disease (e.g. cystoid macular edema, retinal vein occlusion and hypertensive retinopathy); dermatological disease (e.g. melanoma, hair loss and basal cell carcinoma); liver disease (including hepatitis, hepatitis C virus, hepatitis B virus, diabetes, cirrhosis and hepatocellular carcinoma); kidney/renal disease (polycystic kidney disease); and auditory disease (e.g. tinnitus, vertigo, hearing loss and deafness). (I) is useful in therapeutic, veterinary, diagnostic, target validation, genomic discovery, genetic engineering and pharmacogenomic applications. Tests details are described but no results given. ADVANTAGE - (I) has improved: stability; bioavailability; and/or activation of cellular responses mediating RNA interference. (I) minimizes toxicological profile. (I) reduces aggregation over the positively-charged liposome formulations in which the biologically active molecule can be encapsulated in cationic lipids. (I) provides serum stable and reduces the presence of side effects. DESCRIPTION OF DRAWING(S) - The figure shows a lamellar structure and inverted hexagonal structure that can be adopted by a formulated molecular composition.
Manual Codes (1)	A10-E07: Polymerisation, polymer modification -> Chemical modification [others] -> Chemical modification by esterification
	A12-V01: Polymer applications -> Medical, dental, cosmetics and veterinary [others] -> Medicines, pharmaceuticals
	A12-V03C2: Polymer applications -> Medical, dental, cosmetics and veterinary [others] -> Equipment, splints, sutures -> Dental, sterilisation and hygiene; testing, diagnosis and pathology -> Testing, diagnosis, pathology
	A12-W11L: Polymer applications -> Other applications [others] -> Chemical engineering, pollution control -> (Immobilised) enzymes or microorganisms, microbiology (polymer use)
	B01-D02: Steroids -> Steroids - saturated ring'A' -> Androstanes
	B04-B01B: Natural products (or genetically engineered), polymers -> Animal, microbiological and general extracts -> Oils, fats [general] -> Fats, lanolin, lipids
	B04-C03C: Natural products (or genetically engineered), polymers -> Polymers -> Polymers [general] -> Polyethers
	B04-E01: Natural products (or genetically engineered), polymers -> Nucleic acids -> Nucleic acid general and other
	B04-E07C: Natural products (or genetically engineered), polymers -> Nucleic acids -> Other non-coding sequences -> Sirna (short interfering RNA)
	B10-E04D: Aromatics and cycloaliphatics (mono and bicyclic only), aliphatics -> Hydroxy compounds -> Alcohols [general] -> Other alcohols
	B12-K04: Diagnostics and formulation types (therapeutic, pesticidal, herbicidal) -> Diagnostics, respiratory active type -> Diagnosis and testing [general]
	B12-M11Q: Diagnostics and formulation types (therapeutic, pesticidal, herbicidal) -> Formulations type -> Tablets, capsules etc. [general] -> Nanoparticles
	B14-A01: Pharmaceutical activities -> Antimicrobials -> Antibacterial [general]
	B14-A02: Pharmaceutical activities -> Antimicrobials -> Antiviral [general]
	B14-A04: Pharmaceutical activities -> Antimicrobials -> Antifungal general and other

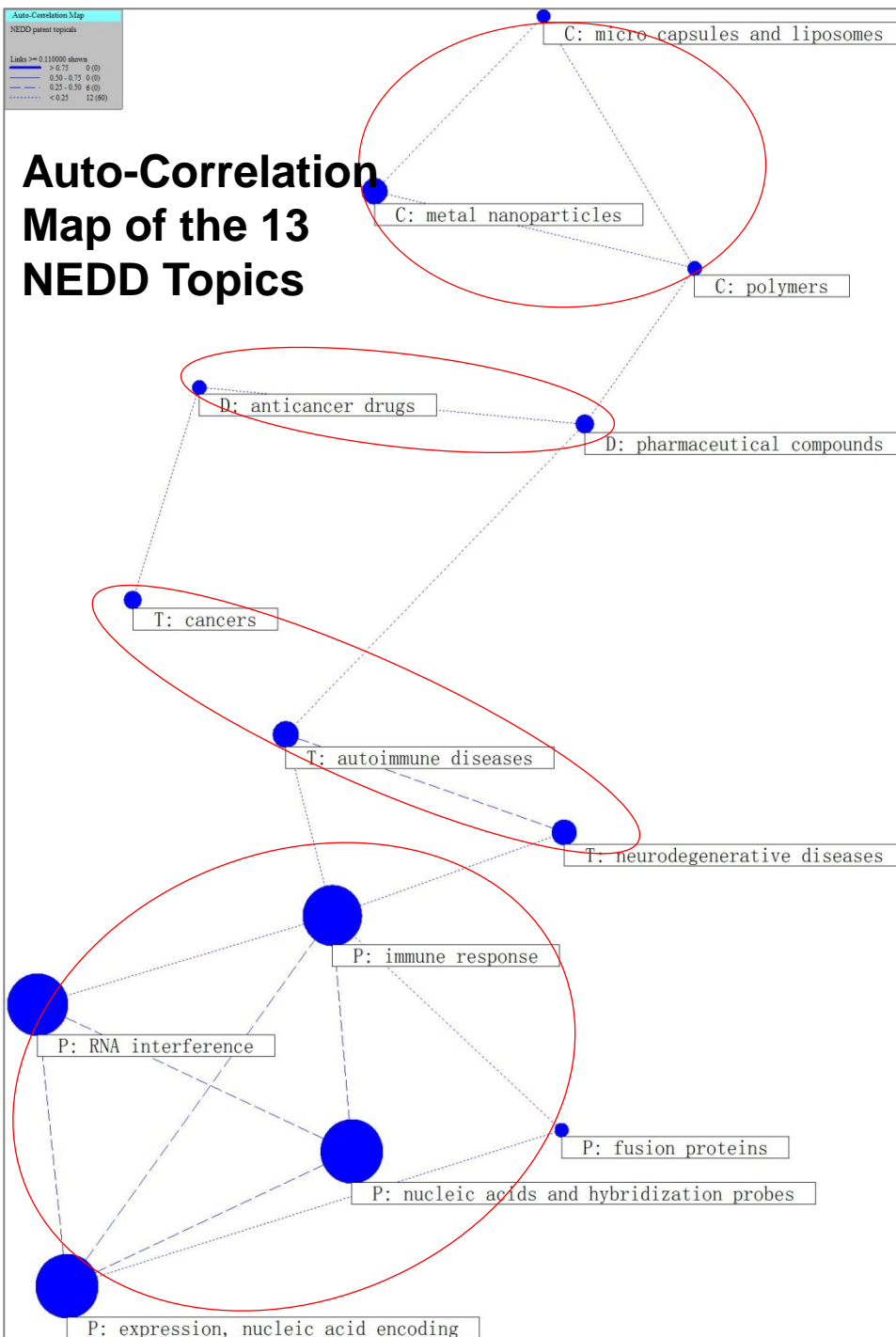
Processes for first two steps



Combined topics in NEDD patents

#	No. of records	Topics	Words or Phrases	Lists
1	3437	P: expression, nucleic acid encoding	expression, polypeptide, polynucleotide, host cell, isolated polypeptide, isolated polynucleotide, inhibitors, recombinant vector, microarrays, polypeptide activity, nucleic acid amplification reaction, gene chips, exogenous polynucleotide, expression dose	List 1
2	3422	P: nucleic acids and hybridization probes	detection, nucleic acid & hybridization probes, nucleic acid hybridization test methods & nucleic acid probes, polarography & enzyme processes, DNA amplification method	List 1 List 2
3	3319	P: RNA interference	RNA, RNA interference, siRNA, sense strand, antisense strand	List 1 List 2 List 3
4	3241	P: immune response	antibody, antigen, T cell, interferon, cytokine, interleukin, macrophage, CD4, CD8	List 3
5	1446	T: autoimmune diseases	multiple sclerosis, Crohn's disease, systemic lupus erythematosus, ulcerative colitis, glomerulonephritis, myasthenia gravis, scleroderma, atopic dermatitis, Addison's disease, Sjogren's syndrome, adult respiratory distress syndrome, uveitis, allergic rhinitis, Grave's disease, spondylitis, Hashimoto's thyroiditis, autoimmune hemolytic anemia, dermatomyositis, polymyositis, urticaria, autoimmune hepatitis, primary biliary cirrhosis	List 1 List 2
6	1433	C: metal nanoparticles	nanoparticle, temperature, colloid, silicon, iron, gold, nanotube, nanostructure, titanium, silver, nanowire	List 3

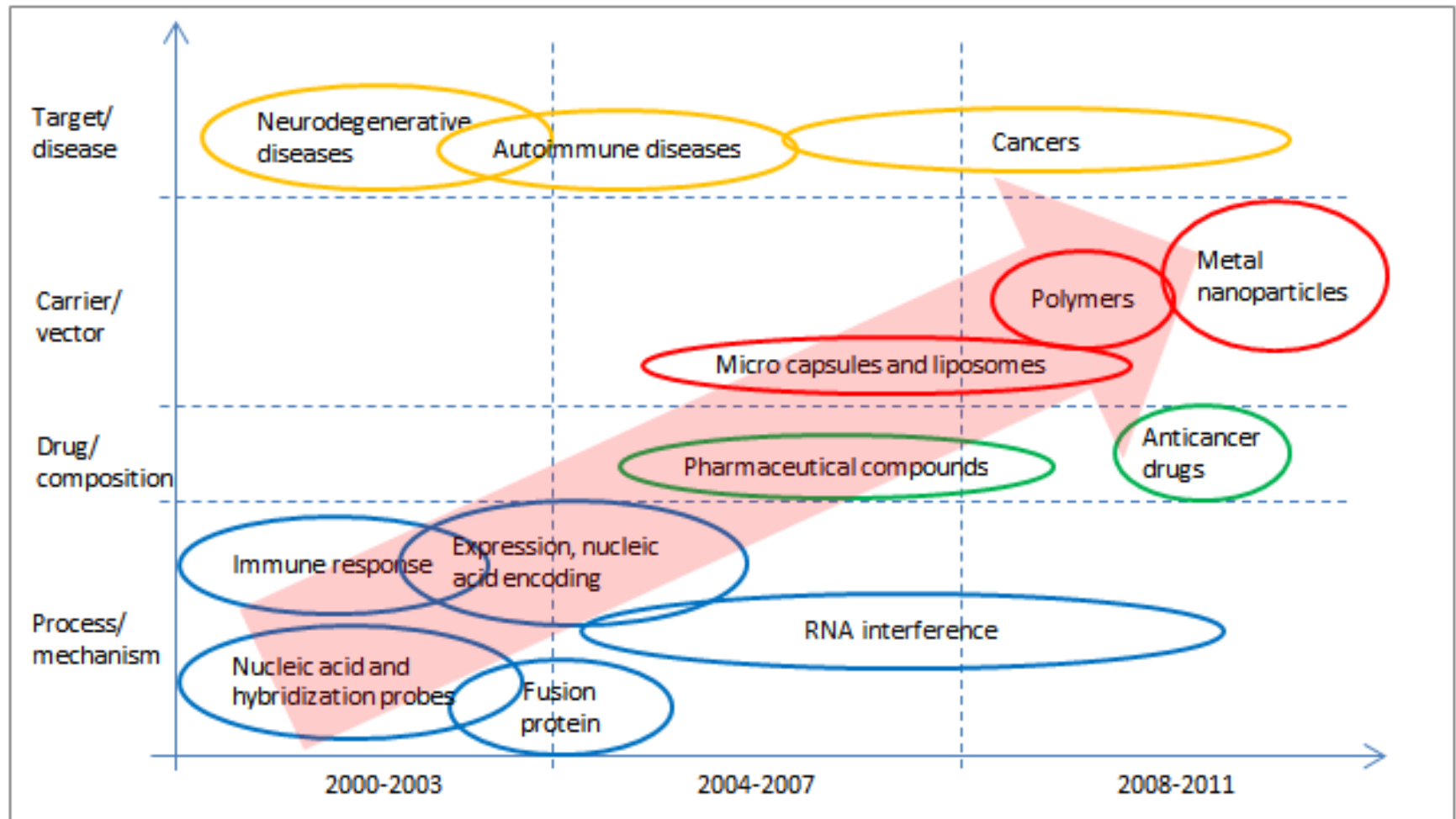
7	1368	T: neurodegenerative diseases	Alzheimer's disease, Parkinson's disease, Huntington's disease, Amyotrophic lateral sclerosis, neurodegenerative diseases, septic shock, spinal cord injury	List 1 List 2
8	1028	D: pharmaceutical compounds	fused ring, mononuclear heterocyclic, polyamines, carboxylic acid & phenol present, carboxylic amides, polycarboxylic acid, aliphatic & cycloaliphatic mono & sulfur acid, amidine & biguanide & guanidine, iso cyanide, hydrazine	List 1 List 2
9	997	T: cancers	breast cancer, prostate cancer, lung cancer, ovarian cancer, pancreatic cancer, colon cancer, liver cancer, cervical cancer, bladder cancer, skin cancer, colorectal cancer, stomach cancer, brain cancer, neck cancer, testicular cancer, esophageal cancer, gastric cancer	List 1
10	676	P: fusion proteins	fusion protein, polypeptide production, fusion genes & transgenes, encoding fusion protein	List 2
11	584	C: polymers	polyethylene, polyethylene glycol, imine, polyethyleneimine, polyether, polyalcohol	List 2 List 3
12	495	C: micro capsules and liposomes	liposome, emulsion, capsule, micelle	List 3
13	341	D: anticancer drugs	doxorubicin, paclitaxel, cisplatin, methotrexate, fluorouracil, camptothecin, dexamethasone	List 1 List 3



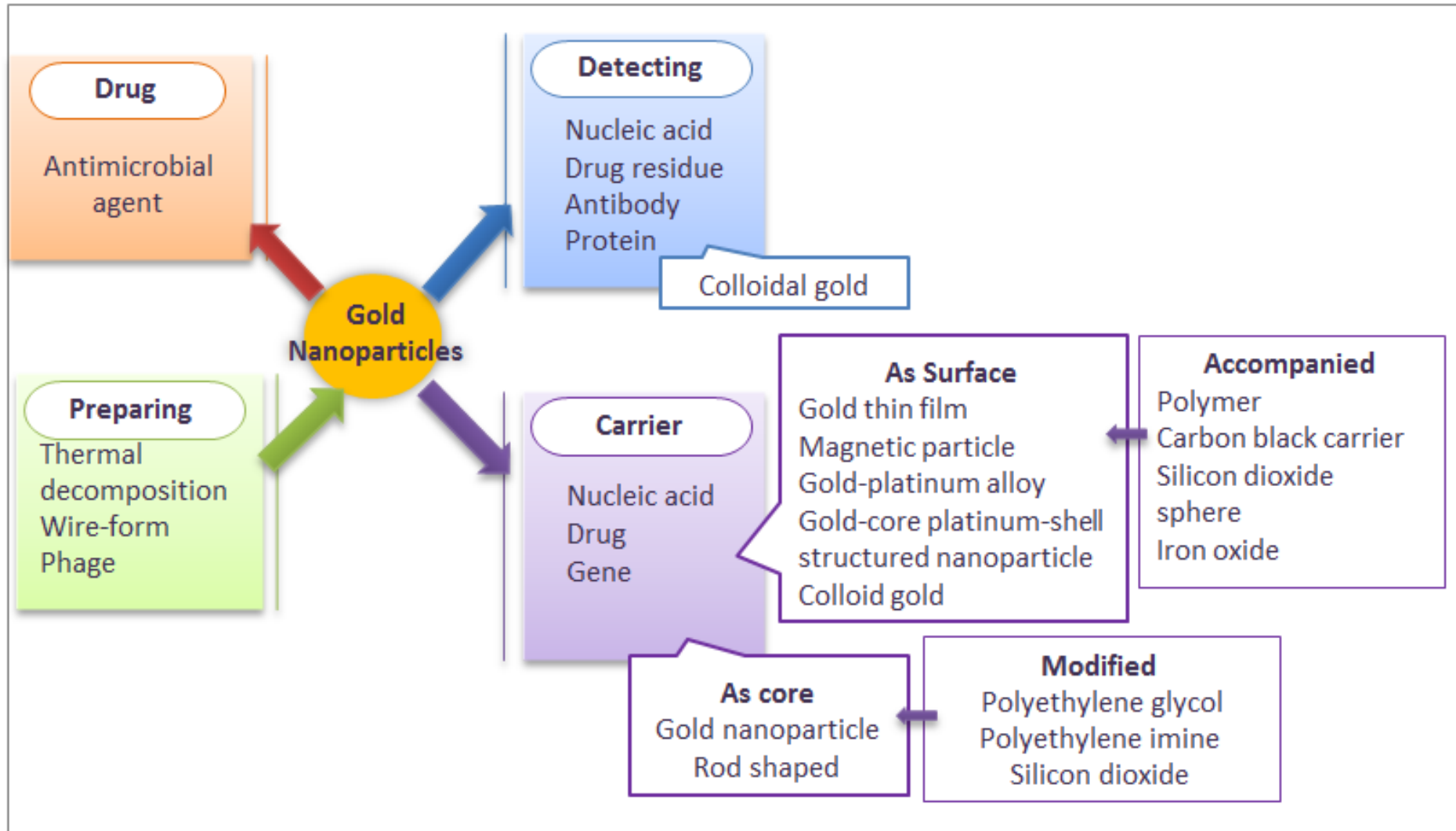
Based on the details of these clusters and their correlations, we divide them into **sub-systems**

- **P: process/mechanism**
 - **D: drug/composition**
 - **C: carrier/vector**
- +
- **T: target/disease.**

Developmental Pathways – Locating the 13 NEDD Topics



Application of Gold Nanoparticles



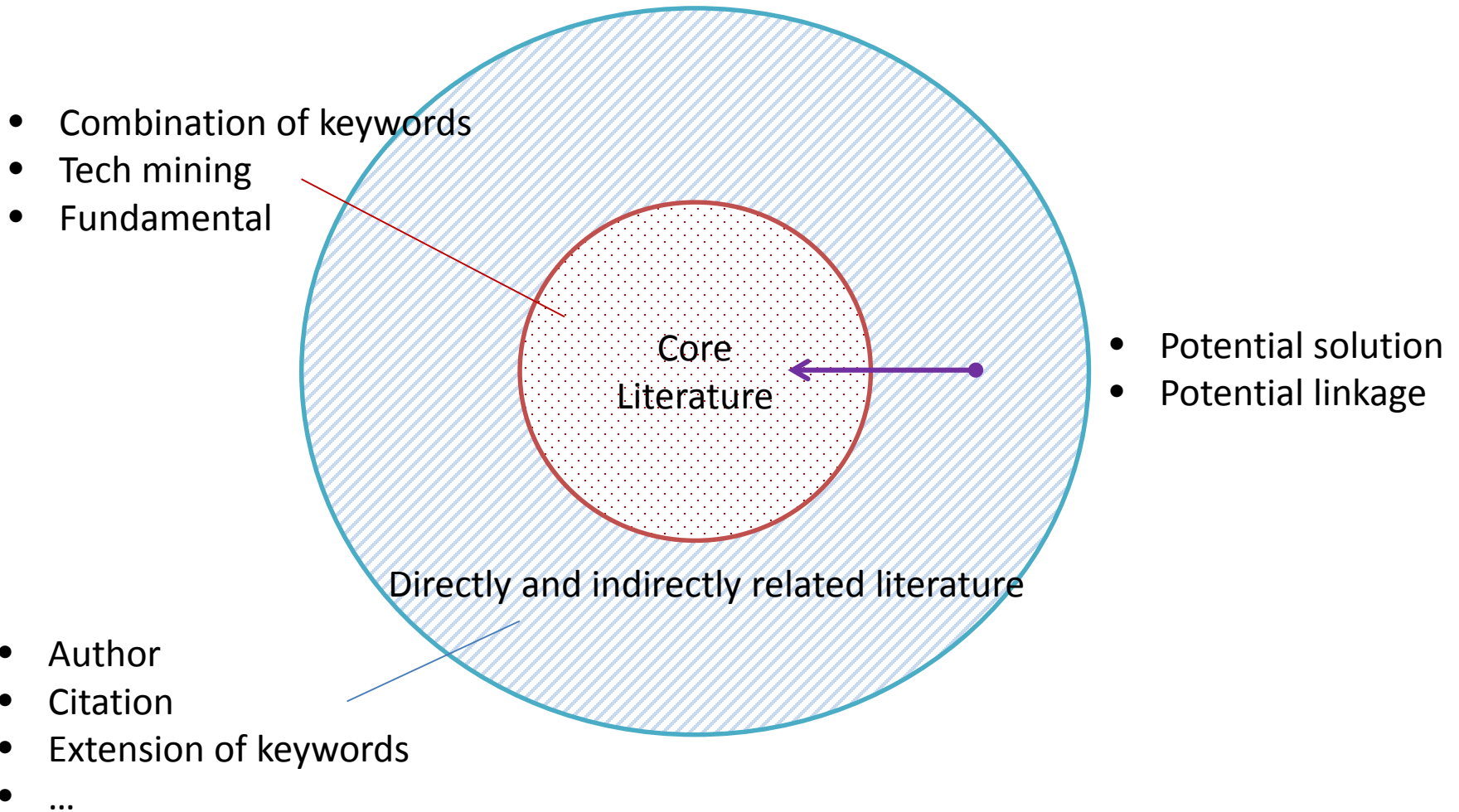
Conclusions

- A systematic process to obtain patent topical information, combining: 3 different topical information sources for patents, software tools for data extracting & cleaning, and specialist review.
- Comparing with traditional, rough topical clustering, the quality and accuracy of our topics are improved.
- Importing external topical information sources, related MeSH terms, enriches our technical intelligence especially novel topics.

Weaknesses of current study

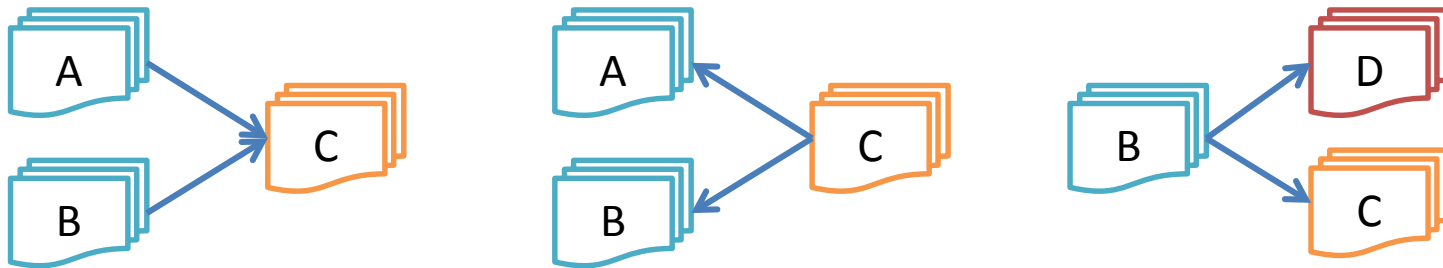
- The linkages and hot topics we analyzed have been reported in literature more or less.
- Technically, they are not potential opportunities.
- Potential concepts/topics should have heretofore not been linked, even radical.
- Start with problem, generate potential solution(s).
- Start with technology, generate potential application(s).

Next step – Generic framework of LRDI (Literature-related Discovery and Innovation)



Next step

- Compare results of different approaches to retrieve directly or indirectly related literature.



- The potential linkage/solution cannot be groundless. Assess to what extent should we expand the query.



- Vetting potential candidates.

Thank You!

Jing Ma

majing881003@163.com