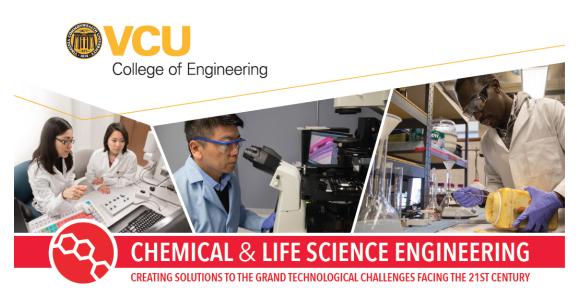# Development of natural language processing tools for nanoparticle design

Nastassja Lewinski, Ph.D.
Department of Chemical and Life Science Engineering

Nano WG – February 28, 2019

1

VCU
College of Engineering

CHEMICAL & LIFE SCIENCE ENGINEERING

CREATING SOLUTIONS TO THE GRAND TECHNOLOGICAL CHALLENGES FACING THE 21ST CENTURY

DEGREE PROGRAMS

B.S.
IN CHEMICAL & LIFE
SCIENCE ENGINEERING

M.S.
IN CHEMICAL & LIFE
SCIENCE ENGINEERING

Ph.D.
IN CHEMICAL & LIFE
SCIENCE ENGINEERING

2017-18 FACULTY ACCOMPLISHMENTS

$7.5 **MILLION**
IN RESEARCH
EXPENDITURES

51 **ISI INDEXED**
REFEREED PUBLICATIONS
PUBLISHED

FACULTY INITIATIVES

medicines for all
INSTITUTE

CERCAS

RESEARCH AREAS

➡ **STEM CELL ENGINEERING**
➡ **NANOTECHNOLOGY**
➡ **PHARMACEUTICAL
ENGINEERING**
➡ **MATERIAL SCIENCE**
➡ **SYSTEMS BIOLOGY**

VCU College of Engineering

Agenda

- Background / Motivation
- Natural language processing
- Case studies
- Recommendation

Gold Nanoshells

Clinical Trials

- NCT00848042 (2009) - Pilot Study of AuroLase(tm) Therapy in Refractory and/or Recurrent **Tumors of the Head and Neck**
- NCT01679470 (2012)- Efficacy Study of AuroLase Therapy in Subjects With Primary and/or **Metastatic Lung Tumors**
- NCT02680535 (2016) - MRI/US Fusion Imaging and Biopsy in Combination With Nanoparticle Directed Focal Therapy for Ablation of **Prostate Tissue**
- NCT02217228 (2014) - Pivotal Study of Sebacia Microparticles in the Treatment of **Acne Vulgaris**
- NCT02758041 (2016) - A Study to Evaluate Sebacia Microparticles in Patients With **Inflammatory Acne Vulgaris**

4

Database of Quantum Dot In Vivo Studies

Table is sortable by heading. Click on heading title (in gray) once to sort by that heading in ascending order, twice for descending.

For a copy of the corresponding Excel sheet and Endnote library, please email nal(a)rice.edu.
Feel free to email any comments/suggestions or if you would like your study added to the database.
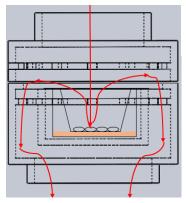
Thank you for visiting!
Link>

Model System	Nanoparticle	Average NP Size	Concentration	Exposure Duration	Results	Author	Year	Reference
BALB/c mice	CdSe/ZnS	N/A, 550 & 625 nm (em)	100-200 mg / 0.1-0.2mL	5, 20 min	coming soon	Akerman	2002	Link
Tadpole embryo	CdSe/ZnS	10-15 nm	2.3 µM	1 hr	coming soon	Dubertret	2002	Link
nude mice	CdSe/ZnS	9, 14 nm	1-20 µM	1 s	coming soon	Larson	2003	Link
Sprague-Dawley rats	CdSe/ZnS	10 nm	2.5 µM	3 hr	coming soon	Lim	2003	Link
BALB/c mice	CdSe/ZnS	N/A, 606, 635, 645, 655nm (em)	50-500 pmol / 50-200 µL	1-3 hr	coming soon	Ballou	2004	Link
Rats	CdSe	N/A, 582 nm (em)	62.5 µg/mL	24 hr	coming soon	Derfus	2004	Link
nude mice	CdSe/ZnS	10-15 nm	20 nM	24 hr	coming soon	Gao	2004	Link
BALB/c mice	CdSe/ZnS	N/A, 520 nm (em)	2 µL of $5x10^7$ cells / mL QD exposed cells	2 hr, 1-7 days	coming soon	Hoshino	2004	Link
nude mice	CdTe/CdSe	15-20 nm	400 pmol	3-4 min	coming soon	Kim	2004	Link
nude mice	QD	N/A, 608 nm (em)	35 µmol	3 s	coming soon	Levene	2004	Link
nude mice	CdSe/ZnS	N/A, 510 nm (em)	10 pmol	5 hr	coming soon	Voura	2004	Link
nude mice	CdSe	4-8 nm	250 pmol	3 hr	coming soon	Ballou	2005	Link

VCU College of Engineering

6

Portable In Vitro Exposure Cassette

- Adapts a 37 mm filter cassette
- Encloses a cell culture insert
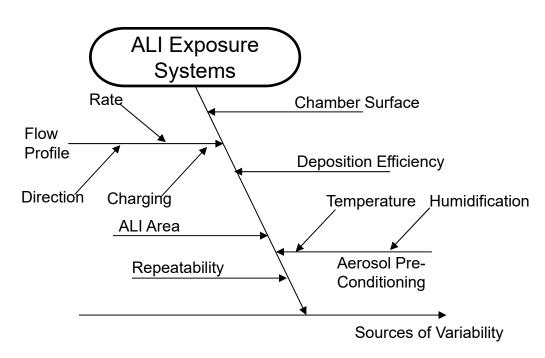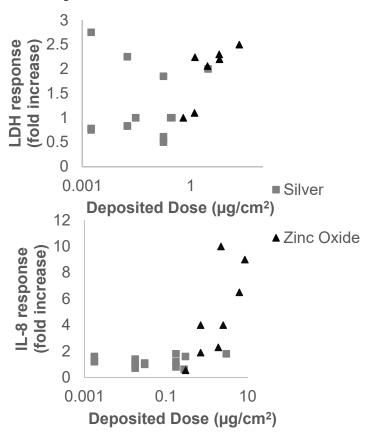- Allows for continuous aerosol flow
- Maintains 37° C, >80% RH

Lewinski N, Secondo L, Ferri J, Process Safety Progress, 2019.
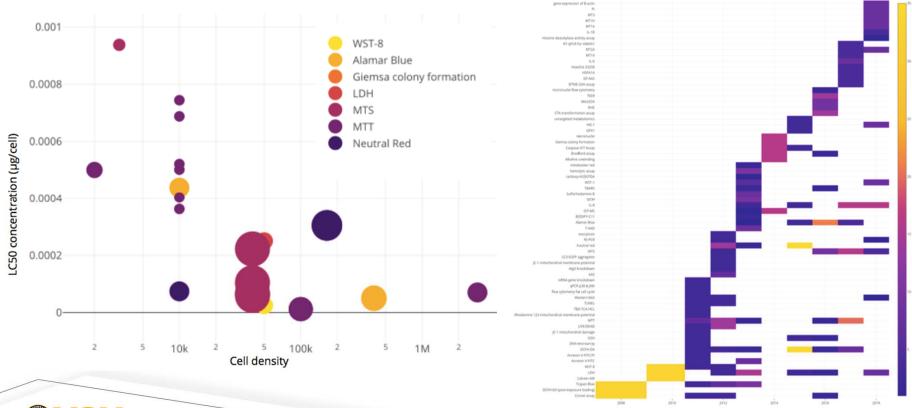Secondo L, Wygal N, Lewinski N, Journal of Visual Experiments, 2019.
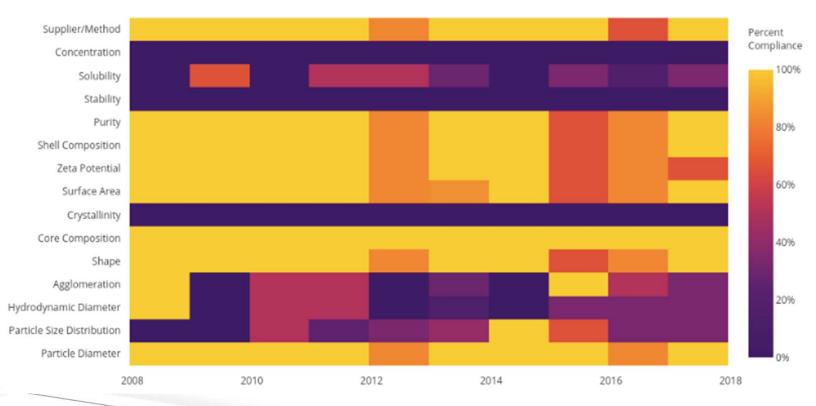
Inter-laboratory comparison

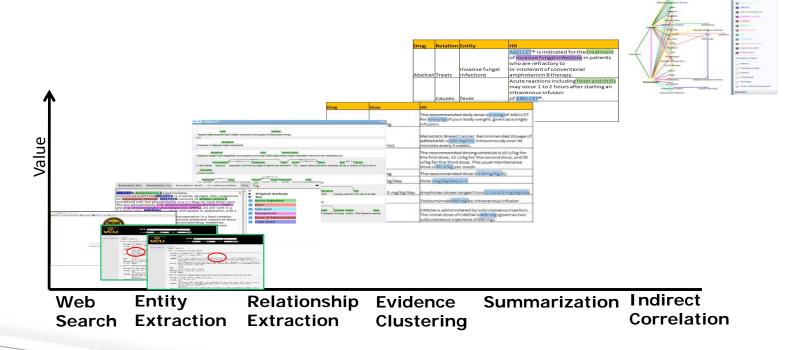Secondo L, Liu N, Lewinski N. Critical Reviews in Toxicology, 2016

Inter-laboratory comparison

Characterization practices

From articles to knowledge

In collaboration with Dr. Bridget McInnes (Computer Science, VCU)

Web Search   Entity Extraction   Relationship Extraction   Evidence Clustering   Summarization   Indirect Correlation

Using natural language processing techniques to inform research on nanotechnology

Nastassja A. Lewinski[1] and Bridget T. McInnes[*2]

Review

Address:
[1]Department of Chemical and Life Science Engineering, Virginia Commonwealth University, Richmond, VA, USA and [2]Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
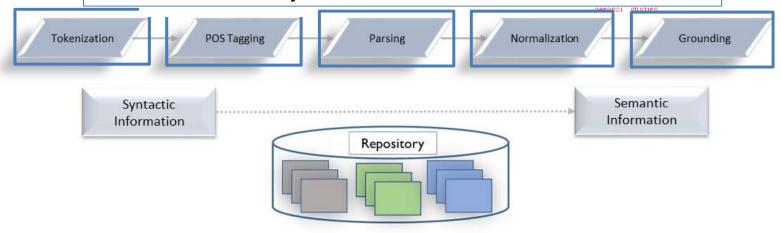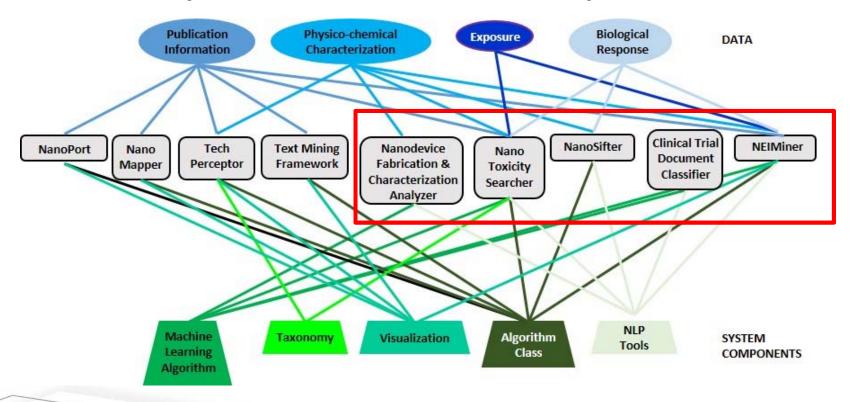
eg. ZnO: **C0043491:Zinc Oxide** nanoparticles: **C1450054:Artificial Nanoparticles** doses: **C0178602:Dosage** over: n/a 1.0: n/a µg/cm^2: **C1631160:Microgram/cm2** were: n/a necessary: n/a to: n/a produce: n/a biological: **C0205460:Biological** responses: **C0871261:Response process** in: n/a aerosol: **C1112870:Aerosol Dose Form** studies: **C0947630:Scientific Study**

Tokenization → POS Tagging → Parsing → Normalization → Grounding

Syntactic Information ⟶ Semantic Information

Repository

Reported nano NLP systems

Lewinski N, McInnes B. *Beilstein Journal of Nanotechnology,* 2015

Named Entity Recognition

| System	Corpora (annotation)	NLP tools	Results
NanoDevEx (2010-2016)	NaDev – 2 articles on nanodevices (Xconc Suite)	YamCha (parser), GPoSSTL (POS tagger)	System F-score = 0.71
NEI Miner (2012-2016)	94 articles from ERDC Environmental Laboratory	ABMiner (NLP toolkit)	System F-score = 0.81
Nano Toxicity Searcher (2013)	654 PubMed abstracts	Abner (NER), Mallet (NLP toolkit)	System F-score = 0.91
NanoSifter (2014)	100 PubMedCentral articles on dendrimers (GATE)	ANNIE (GATE module)	System F-score = 0.91

VCU College of Engineering
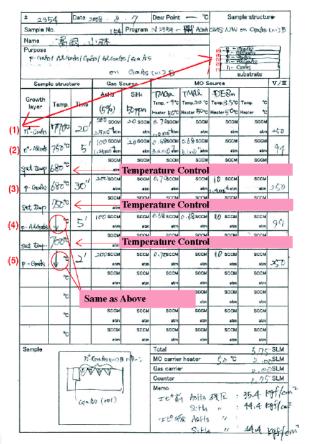
Figure 1: An example of a SA-MOVPE growth parameter record sheet
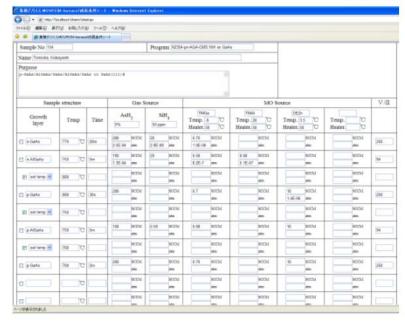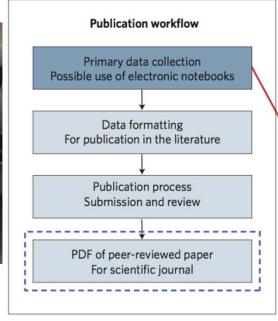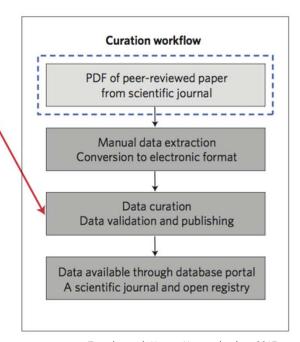
Figure 3: Screenshot of the SA-MOVPE experiment record management system

Yoshioka et al. *Proceedings of 12th Int. Conf. Info. Integr. Web Appl. Serv.* 2010
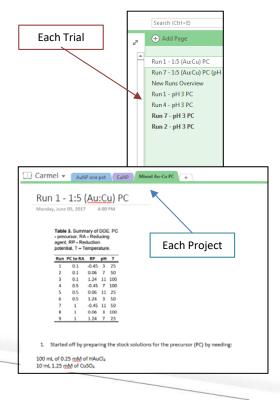
Automation

Chan et al. *Nano Letters* 2010

Tropsha et al. *Nature Nanotechnology* 2017
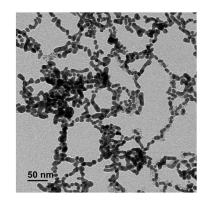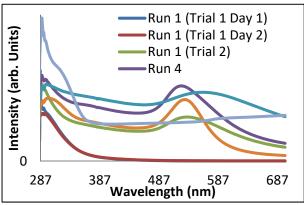
Nanoparticle Synthesis

- Run experiments and record procedure and results in Microsoft OneNote

Each Trial

Each Project

50 nm

ORIGINAL RESEARCH

An annotated corpus with nanomedicine and pharmacokinetic parameters
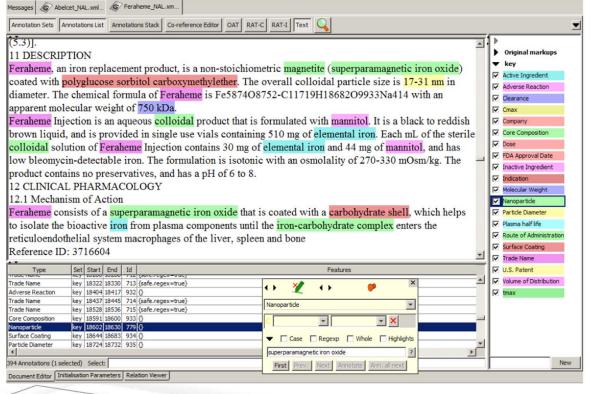
Nastassja A Lewinski[1]
Ivan Jimenez[1]
Bridget T McInnes[2]

[1]Department of Chemical and Life Science Engineering, Virginia Commonwealth University, Richmond, VA, [2]Department of Computer Science, Virginia Commonwealth

Abstract: A vast amount of data on nanomedicines is being generated and published, and natural language processing (NLP) approaches can automate the extraction of unstructured text-based data. Annotated corpora are a key resource for NLP and information extraction methods which employ machine learning. Although corpora are available for pharmaceuticals, resources for nanomedicines and nanotechnology are still limited. To foster nanotechnology text mining (NanoNLP) efforts, we have constructed a corpus of annotated drug product inserts taken from the US Food and Drug Administration's Drugs@FDA online database. In this work, we present

19

Engineered Nanomedicine Database (END)
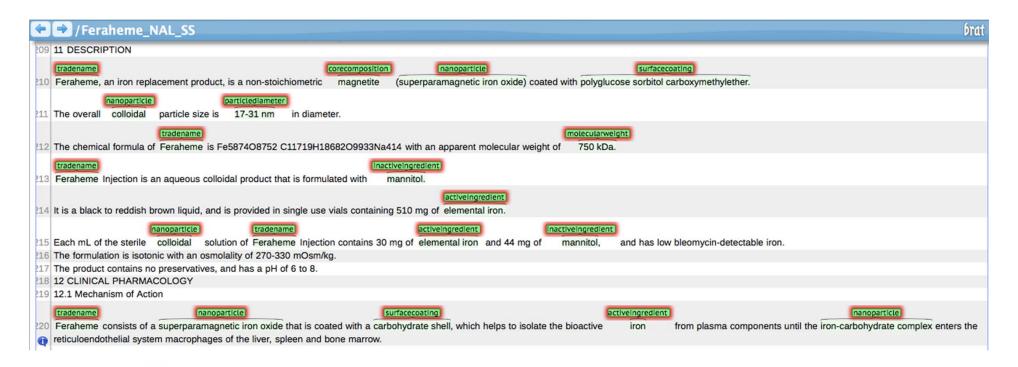

GATE Features:

- Runs on Windows and MacOS

- Off the shelf tools (e.g. ANNIE) built-in

Disadvantages

- Differences in files between Windows and MacOS

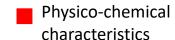- Relationship annotations only by using a taxonomy

VCU College of Engineering

brat Example Annotated Document

Engineered Nanomedicine Database (END)

# instances	entity	# instances	entity	# instances	entity
2161	active ingredient	47	auc	29	volume of distribution
89	core composition	49	clearance	2283	dose
50	molecular weight	41	cmax	1192	route of administration
854	nanoparticle	65	elimination half life	6689	adverse reaction
7	particle diameter	56	plasma half life	1396	indication
62	surface coating	30	tmax		

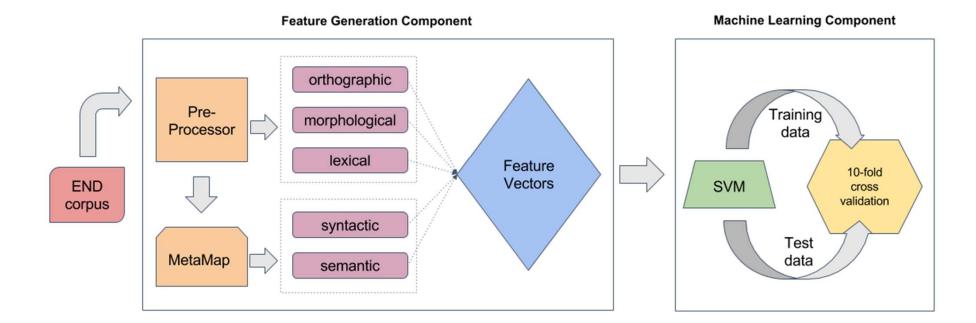■ Physico-chemical characteristics   ■ Pharmacokinetic parameters   ■ Exposure   ■ Biological response
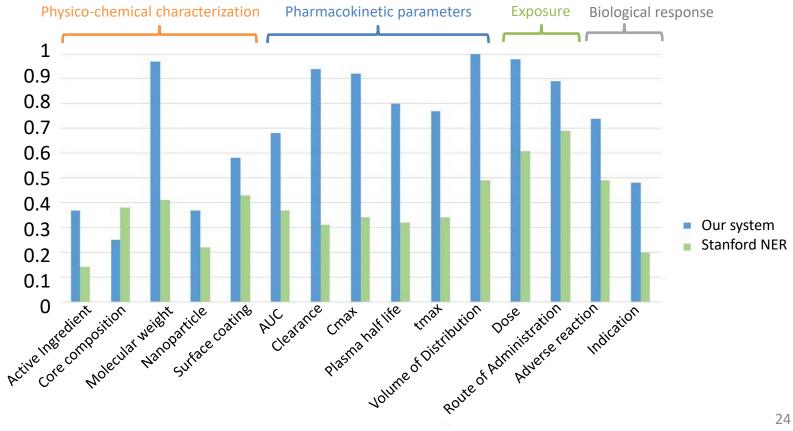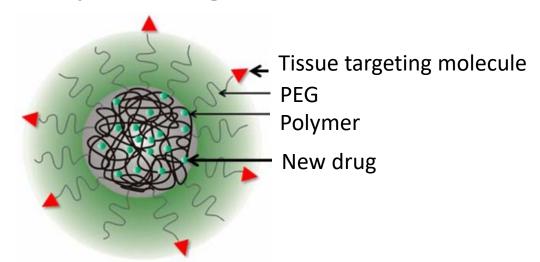
Nano-entity recognition

Evaluation

Nanomedicine Design

- Scenario 1 – synthesis planning

Given information:

New small molecule drug

Desired target tissue

Tissue targeting molecule
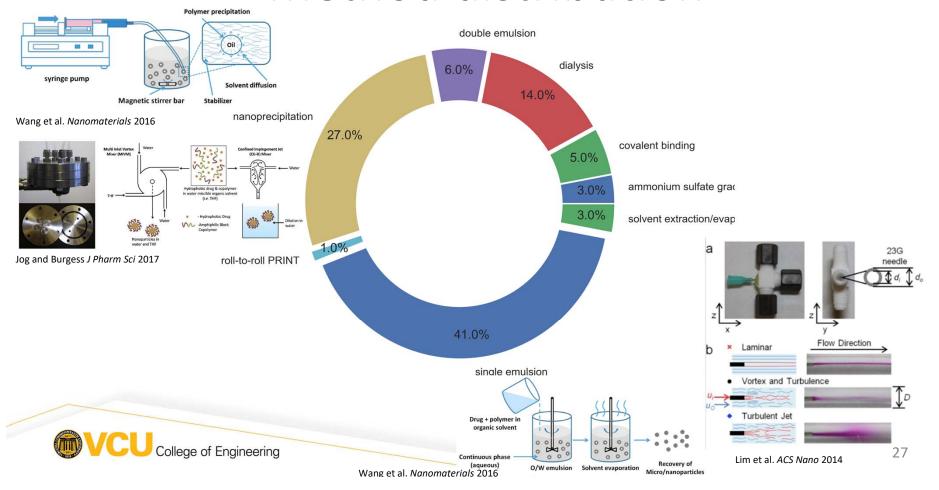PEG
Polymer
New drug

API subset

DOX logP = 1.27   DOC logP = 2.4   PAC logP = 3
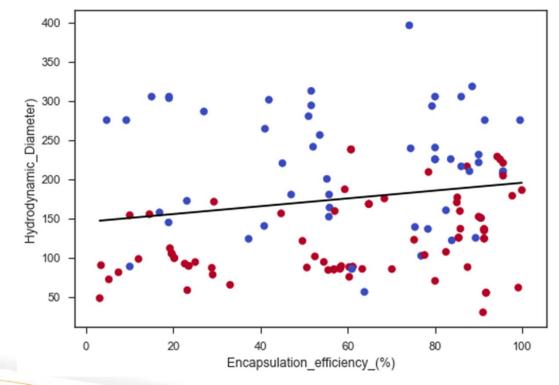
82 papers – 172 synthesis methods
168 batch, 4 flow

Method distribution

Wang et al. *Nanomaterials* 2016

Jog and Burgess *J Pharm Sci* 2017

double emulsion

6.0%

dialysis

14.0%

covalent binding

5.0%

ammonium sulfate grad

3.0%

solvent extraction/evap

3.0%

nanoprecipitation

27.0%

roll-to-roll PRINT

1.0%

single emulsion

41.0%

Wang et al. *Nanomaterials* 2016

Lim et al. *ACS Nano* 2014

VCU College of Engineering
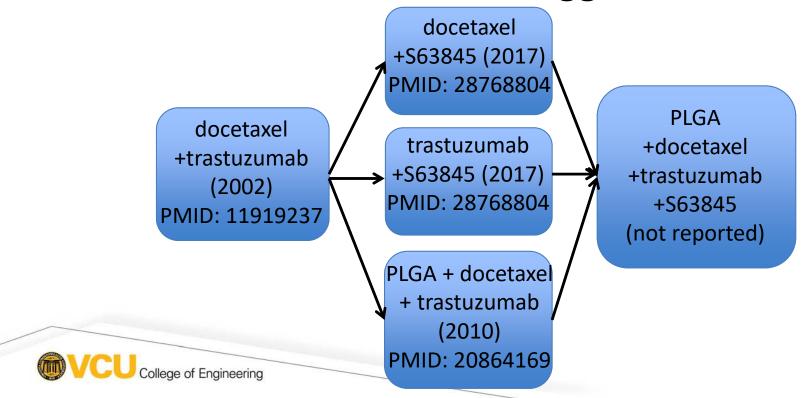
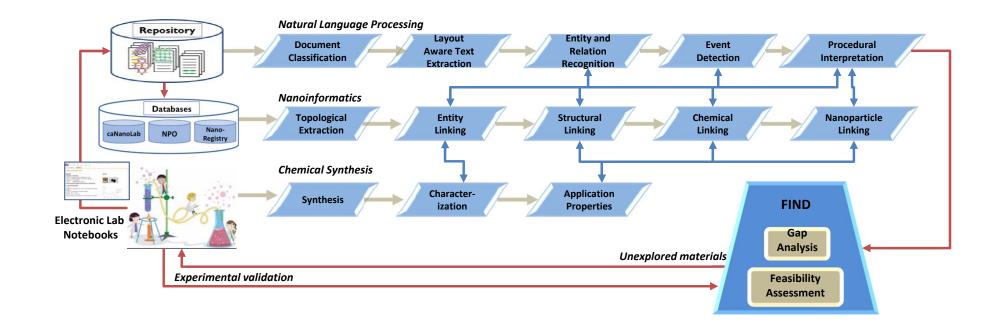# Preliminary findings

Blue = PLGA
Red = PLGA-PEG

Nanomedicine Design

- Scenario 2 – formulation suggestion

Framework for Intelligent Nanotechnology Discovery

Recommendation

- Nanotechnology shared task or challenge
  - BioCreative
  - Text Analysis Conference (TAC)
  - National NLP Clinical Challenge (N2C2)

Acknowledgements

Students

Ivan Jimenez

Tanin Izadi

Marley Hodson

Jacob Jaminet

Robert Kalish

Carly Hughes

Collaborators

Bridget McInnes

Funding

Quest for Distinction

VCU Presidential Research Quest Fund

Nanoinformatics VIP Team December 2018

VCU College of Engineering