HEALTHY BIRTH,
GROWTH & DEVELOPMENT
knowledge integration

Discovering Undiscovered Public Knowledge with Influence Search

Mihai Surdeanu

October 2, 2017

HBGDki

Conflict of interest disclosure

M. Surdeanu discloses a financial interest in Lum.ai. This interest has been disclosed to the University of Arizona Institutional Review Committee and is being managed in accordance with its conflict of interest policies.

Knowledge can be public, yet undiscovered, if independently created fragments are logically related but never retrieved, brought together, and interpreted.

–Swanson, 1986

HBGD*ki*

Rate of publication

- There's a lot of research out there
- 90% of research is never cited[1]
- The readability of literature decreases over time[2]

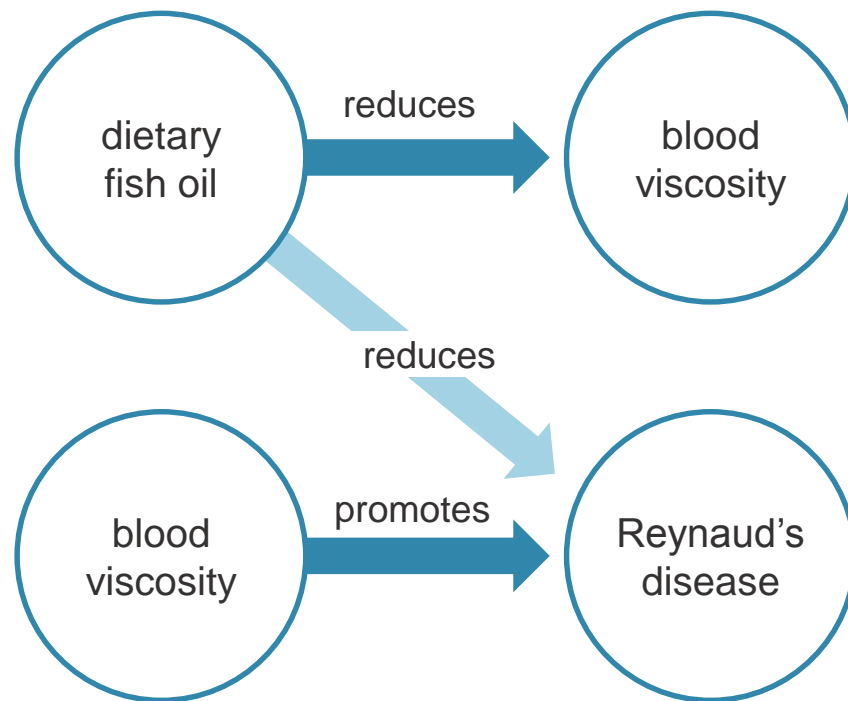Publications indexed by PubMed each year since 1995
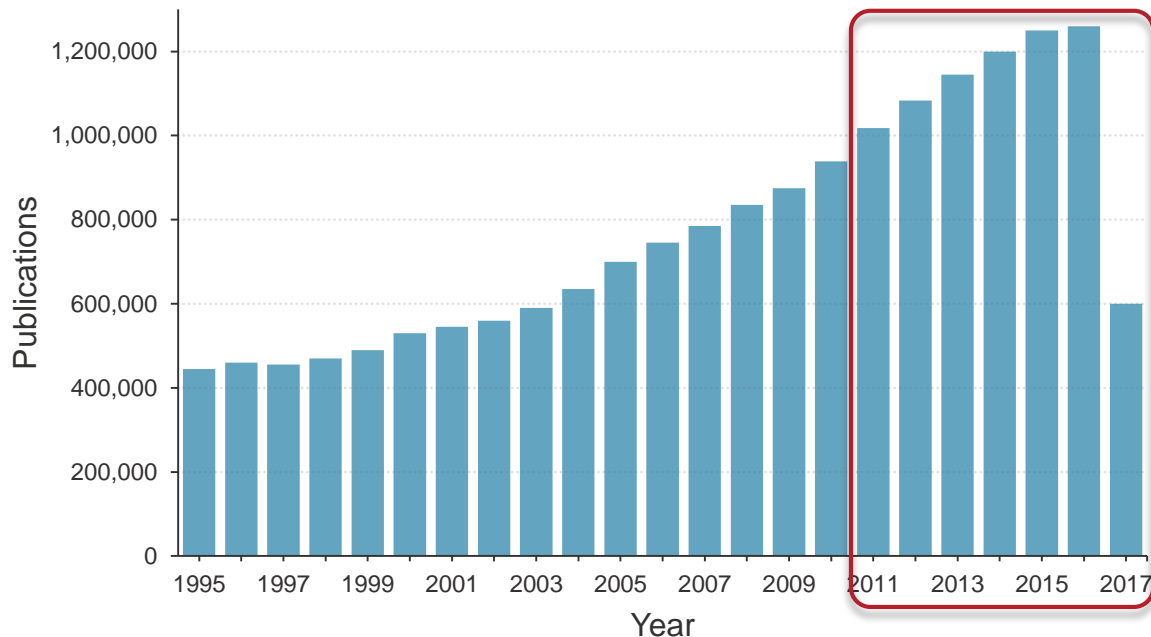
[1] Eveleth R. Academics Write Papers Arguing Over How Many People Read (And Cite) Their Papers. Smithsonian website
  http://www.smithsonianmag.com/smart-news/half-academic-studies-are-never-read-more-three-people-180950222/. Accessed Sep. 20, 2017.
[2] Plaven-Sigray, et al. Research: The readability of scientific texts is decreasing over time. eLife Sciences website 2017;6:e27725 DOI: 10.7554/eLife.27725.
  Accessed Sep. 20, 2017.

Missed opportunities

- There are many independent yet related pieces of knowledge that are not being brought together to form new discoveries.

What if the necessary discoveries to solve problem X have already been made, but these puzzle pieces are scattered across the literature?

dietary fish oil → **reduces** → blood viscosity

dietary fish oil → **reduces** → Reynaud's disease

blood viscosity → **promotes** → Reynaud's disease

Swanson D. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspectives in biology and medicine 30(1), pp. 7 – 18, 1986.
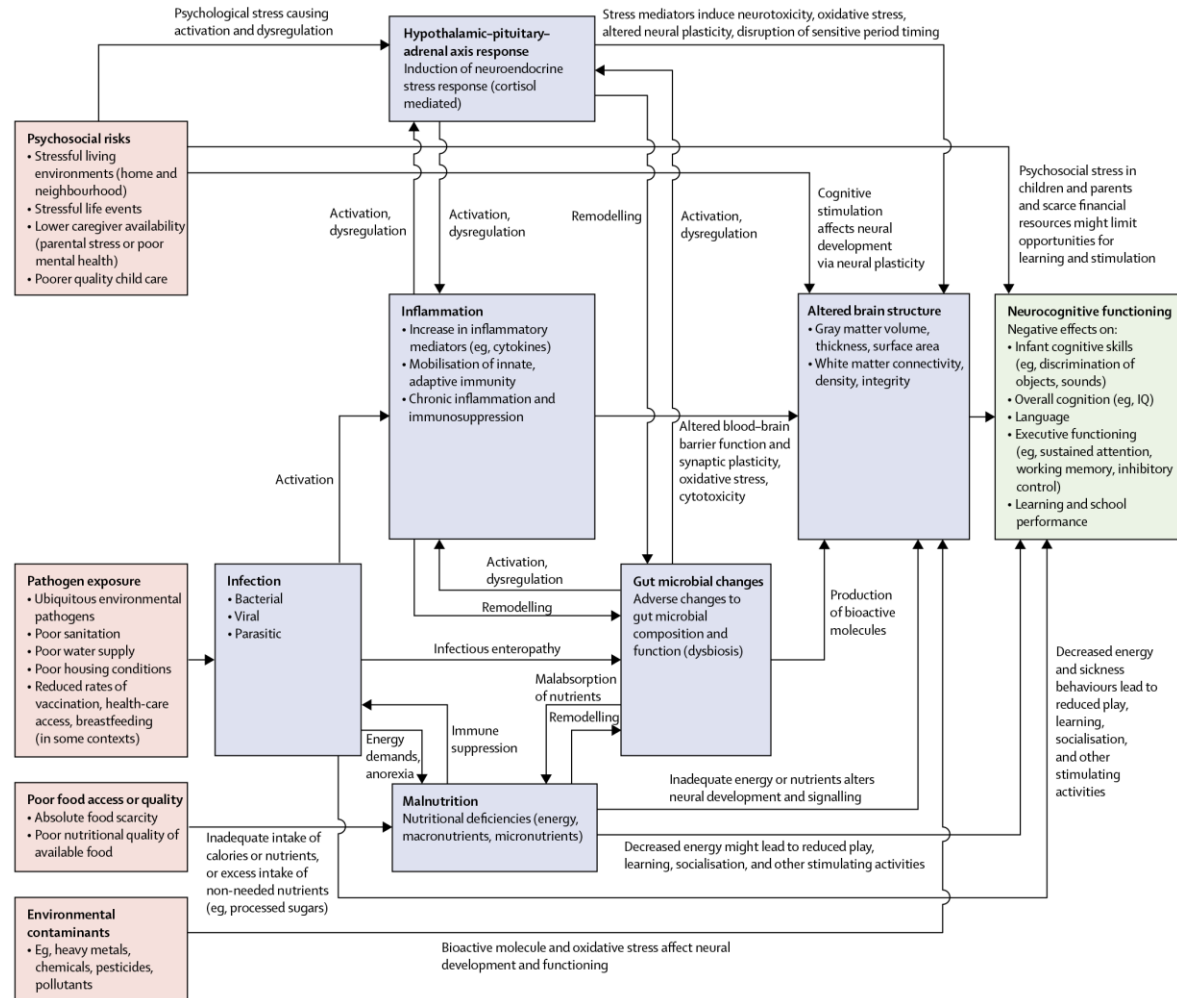
What if the necessary discoveries to solve problem X have already been made, but these puzzle pieces are scattered across the literature?

Psychological stress causing activation and dysregulation

Stress mediators induce neurotoxicity, oxidative stress, altered neural plasticity, disruption of sensitive period timing

Hypothalamic–pituitary–adrenal axis response
Induction of neuroendocrine stress response (cortisol mediated)

Psychosocial risks
- Stressful living environments (home and neighbourhood)
- Stressful life events
- Lower caregiver availability (parental stress or poor mental health)
- Poorer quality child care

Activation, dysregulation

Activation, dysregulation

Remodelling

Activation, dysregulation

Cognitive stimulation affects neural development via neural plasticity

Psychosocial stress in children and parents and scarce financial resources might limit opportunities for learning and stimulation

Inflammation
- Increase in inflammatory mediators (eg, cytokines)
- Mobilisation of innate, adaptive immunity
- Chronic inflammation and immunosuppression

Altered brain structure
- Gray matter volume, thickness, surface area
- White matter connectivity, density, integrity

Neurocognitive functioning
Negative effects on:
- Infant cognitive skills (eg, discrimination of objects, sounds)
- Overall cognition (eg, IQ)
- Language
- Executive functioning (eg, sustained attention, working memory, inhibitory control)
- Learning and school performance

Altered blood–brain barrier function and synaptic plasticity, oxidative stress, cytotoxicity

Activation

Pathogen exposure
- Ubiquitous environmental pathogens
- Poor sanitation
- Poor water supply
- Poor housing conditions
- Reduced rates of vaccination, health-care access, breastfeeding (in some contexts)

Infection
- Bacterial
- Viral
- Parasitic

Activation, dysregulation

Remodelling

Gut microbial changes
Adverse changes to gut microbial composition and function (dysbiosis)

Production of bioactive molecules

Decreased energy and sickness behaviours lead to reduced play, learning, socialisation, and other stimulating activities

Infectious enteropathy

Malabsorption of nutrients

Remodelling

Energy demands, anorexia

Immune suppression

Inadequate energy or nutrients alters neural development and signalling

Poor food access or quality
- Absolute food scarcity
- Poor nutritional quality of available food

Malnutrition
Nutritional deficiencies (energy, macronutrients, micronutrients)

Inadequate intake of calories or nutrients, or excess intake of non-needed nutrients (eg, processed sugars)

Decreased energy might lead to reduced play, learning, socialisation, and other stimulating activities

Environmental contaminants
- Eg, heavy metals, chemicals, pesticides, pollutants

Bioactive molecule and oxidative stress affect neural development and functioning
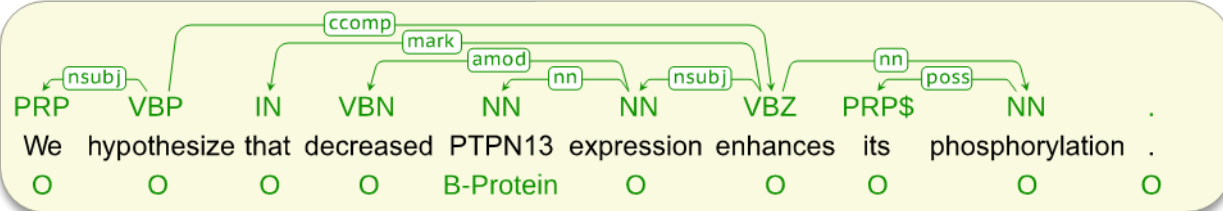
HBGD*ki*

Machine reading and assembly overview

Influence mechanism is a tool that helps make these new discoveries by applying machine reading to very large collections of literature
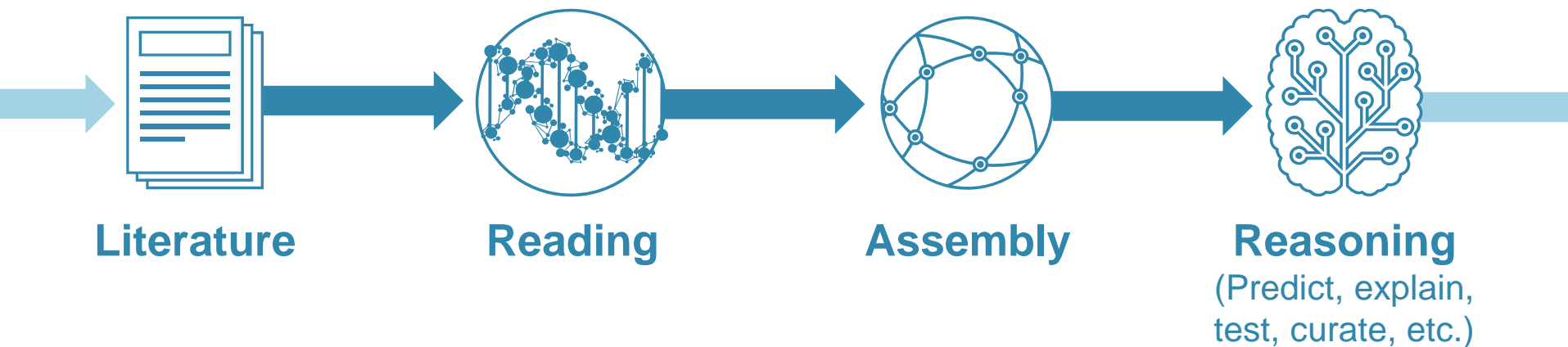
Literature   **Reading**   **Assembly**   **Reasoning**
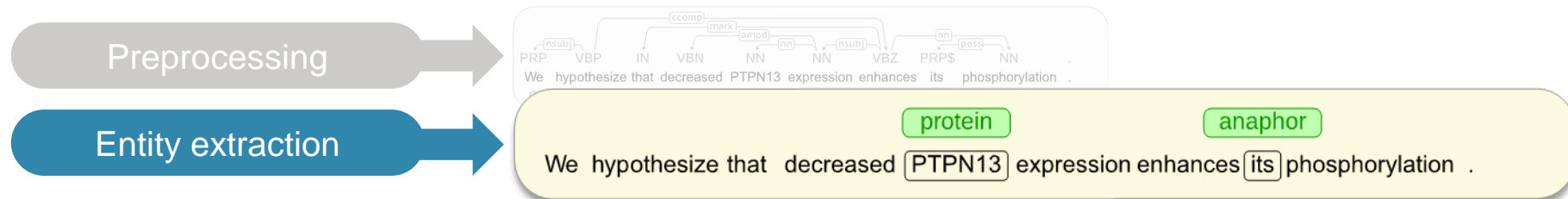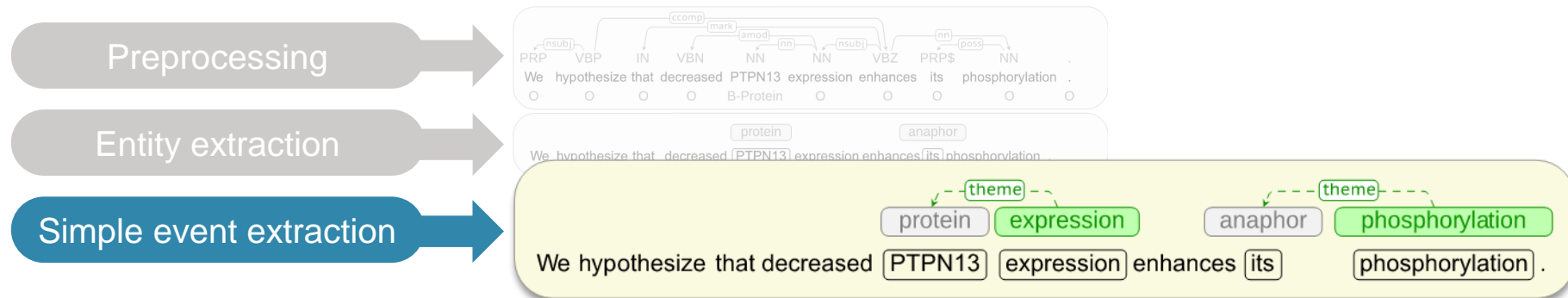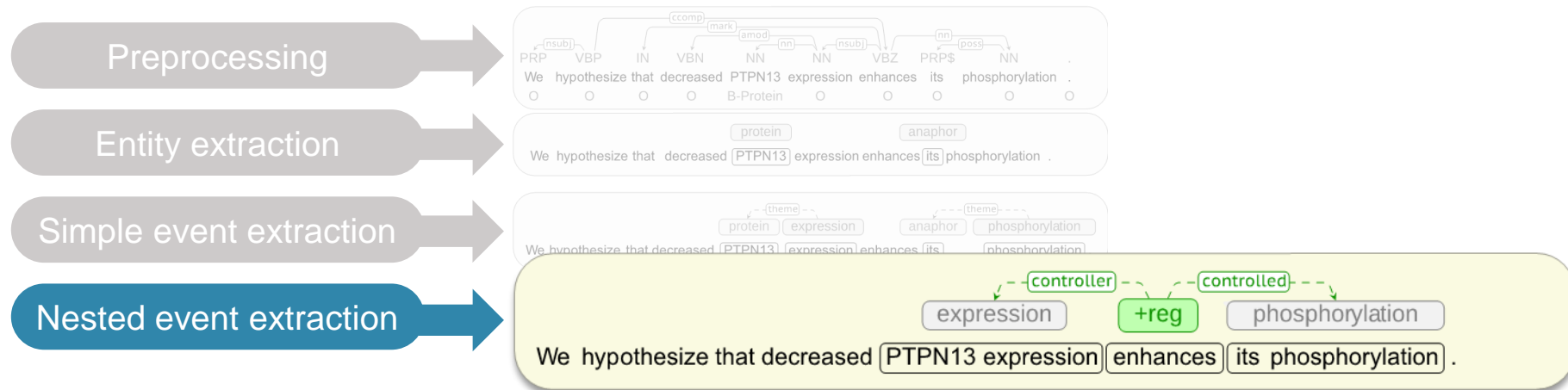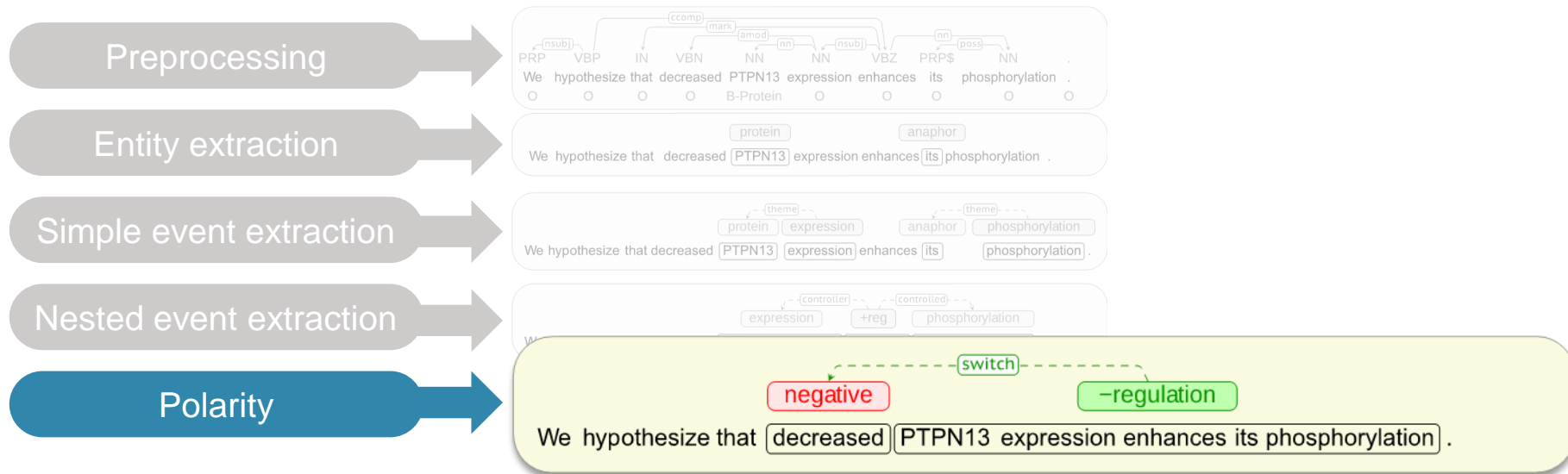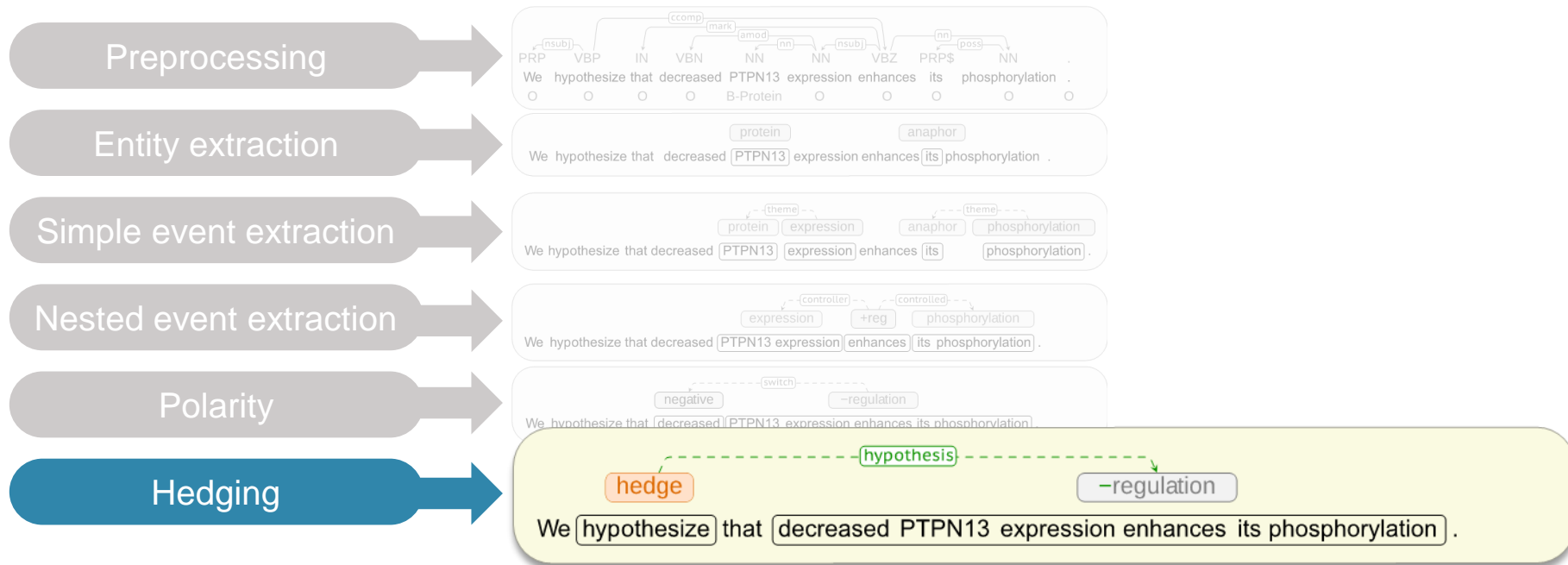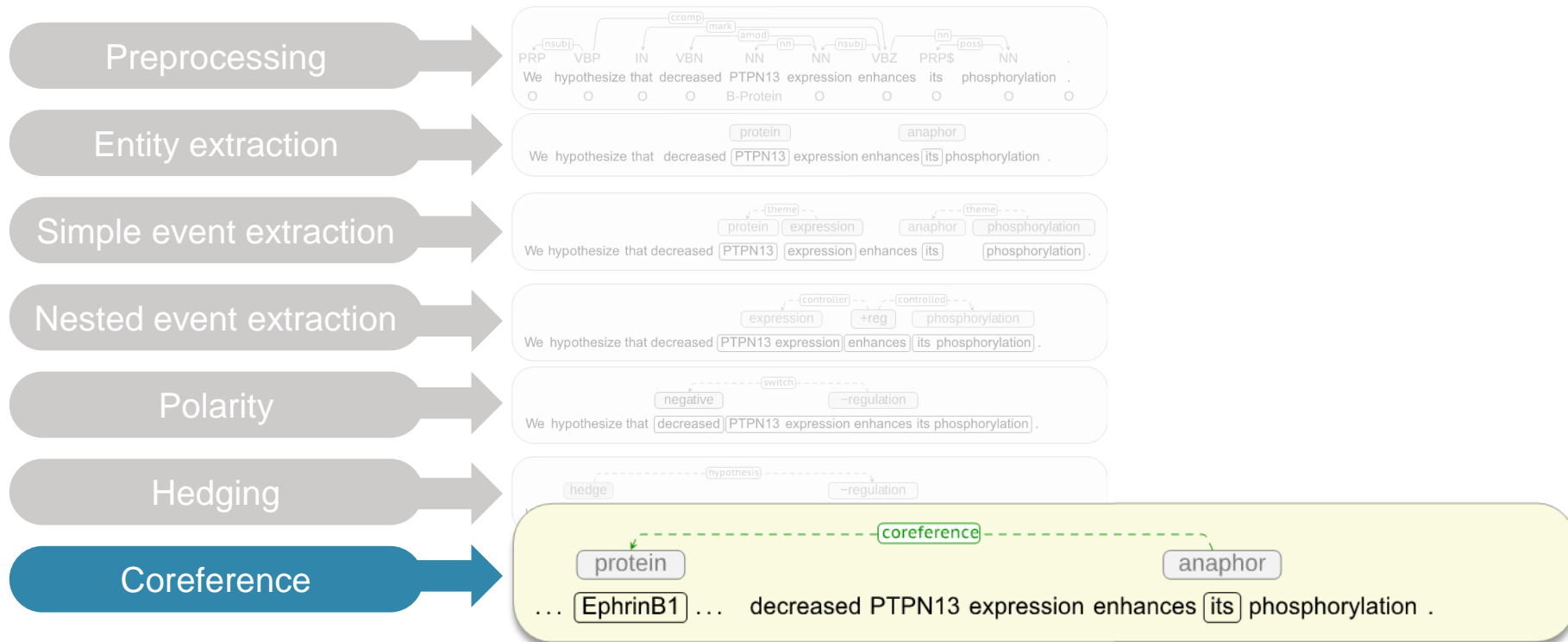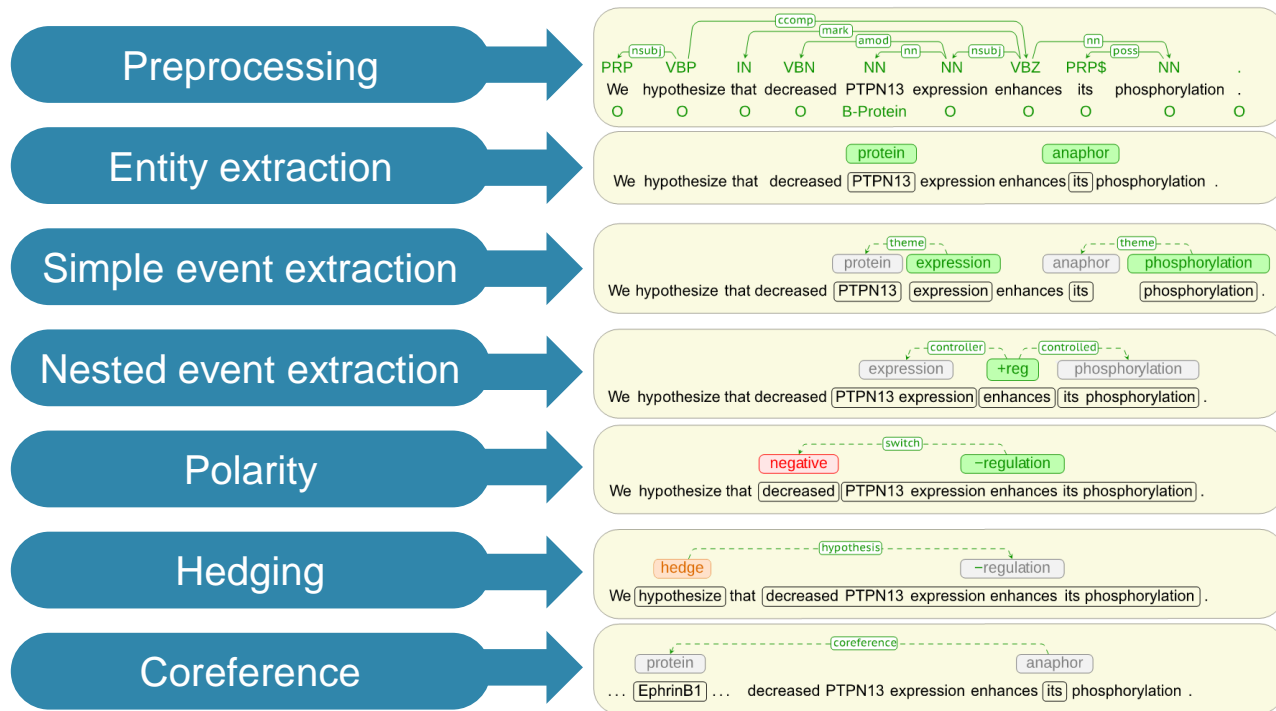(Predict, explain, test, curate, etc.)

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

Preprocessing

Machine reading: how it works

## REACH (REading and Assembling Contextual and Holistic mechanisms)

Preprocessing

Entity extraction

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

Preprocessing

Entity extraction

Simple event extraction

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

Preprocessing

Entity extraction

Simple event extraction

Nested event extraction

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

Preprocessing

Entity extraction

Simple event extraction

Nested event extraction

Polarity

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

- Preprocessing
- Entity extraction
- Simple event extraction
- Nested event extraction
- Polarity
- **Hedging**

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

Machine reading: how it works

REACH (REading and Assembling Contextual and Holistic mechanisms)

Preprocessing

Entity extraction

Simple event extraction

Nested event extraction

Polarity

Hedging

Coreference

Machine assembly: how it works

- Aggregation of findings is crucial to break the "silo nature of science"
- Performed using taxonomies (e.g., taxonomy of proteins) or using linguistic transformations, when such taxonomies do not exist
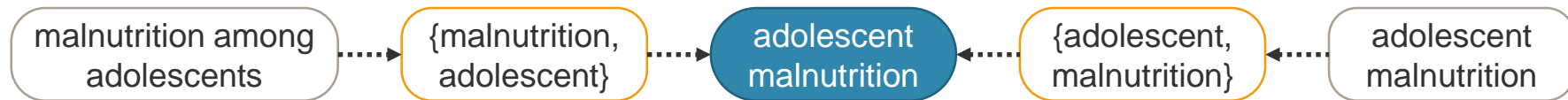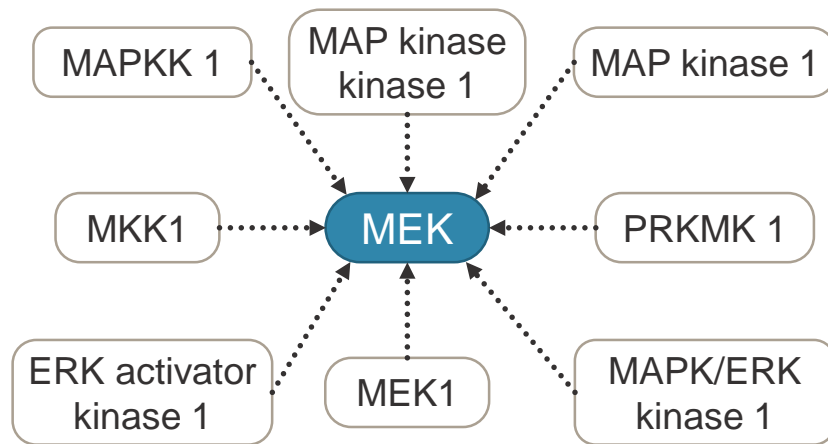
Machine reading and assembly summary

Works across multiple domains and works with and without ontologies

Literature   **Reading**   **Assembly**   **Reasoning**
(Predict, explain, test, curate, etc.)
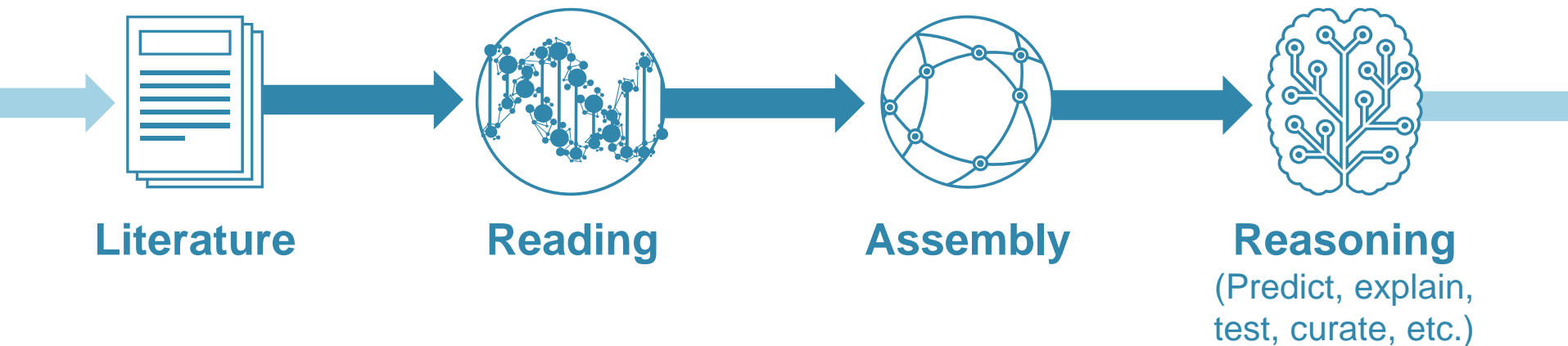
- Search for influence patterns rather than keywords
- Breaks the "silo nature of science" by aggregating findings across multiple domains
- Encourages sharing of models and collaborative model building

Publications indexed

115K PubMed Open Access full papers

- Relevant to children malnutrition and cognitive development
- +2M distinct entities
- +1.8M promotions, 0.8M inhibitions

+26M PubMed abstracts (all Open Access)

- +27M distinct entities
- +16M promotions, 8.5M inhibitions

In the future:

- Publications behind pay walls
- Emails, slides, etc.
- Implied knowledge

vitamin D | 1 | Effect | Search | ⚙ Options

Context

Search boxes used to specify the desired **cause** and/or **effect**, and the maximum number of hops connecting them.

Graph

Copy | CSV | Excel | Filter results:

Filter... | Filter... | Filter...

	Relation	Effect	Evidence
n D	decreases	rickets	seen 12 times
n D	increases	calcium absorption	seen 10 times
n D	increases	intestinal calcium absorption	seen 10 times

Showing 1 to 50 of 3,089 entries

Previous | 1 | 2 | 3 | 4 | 5 | ... | 62 | Next

Show 50 entries

Evidence ✕

vitamin D *decreases* rickets

PMC4160567

. . . Adequate vitamin D is essential to prevent rickets , but adequate calcium is equally important ; if either calcium or vitamin D is deficient , bone health suffers . . . .
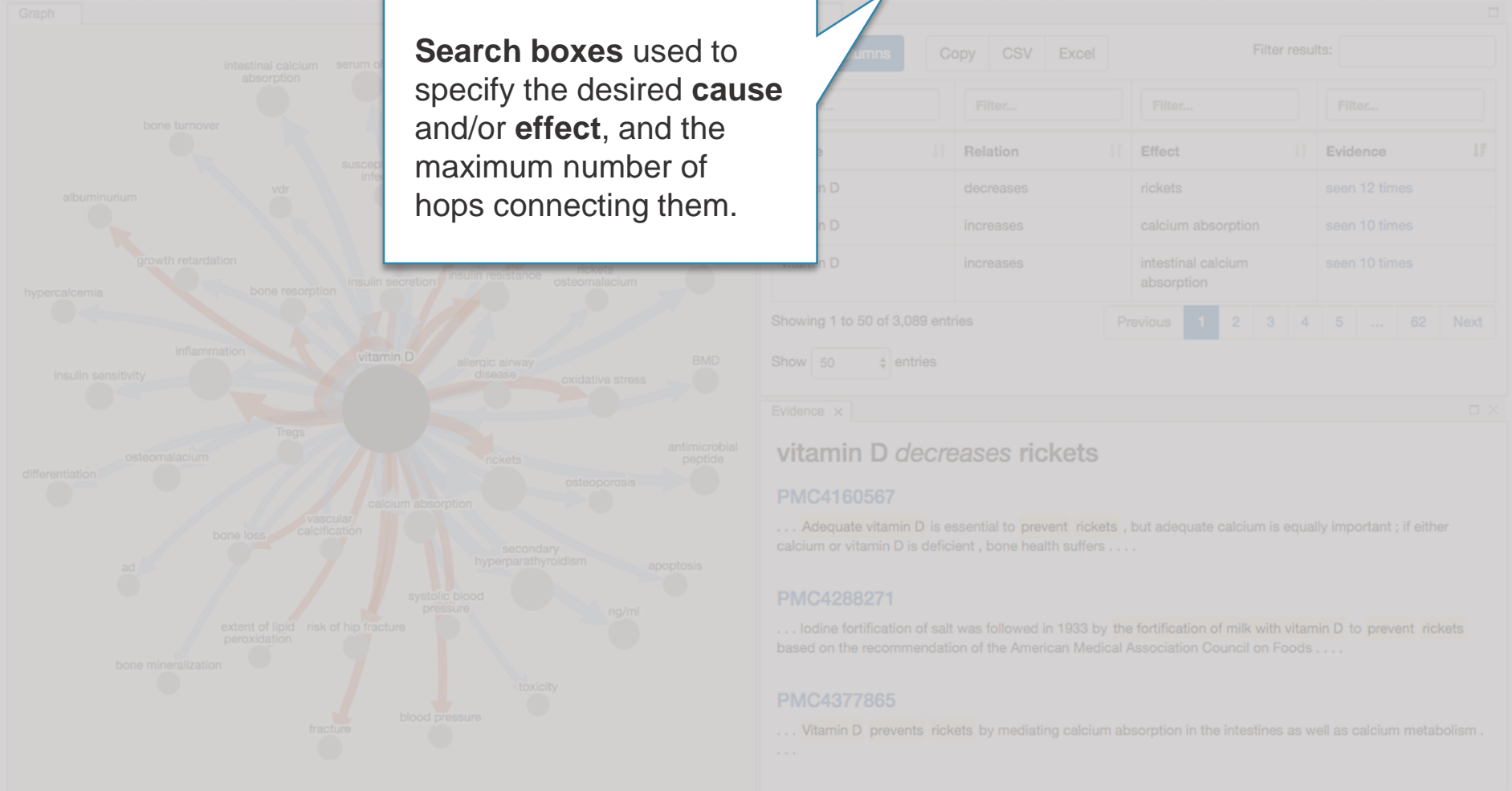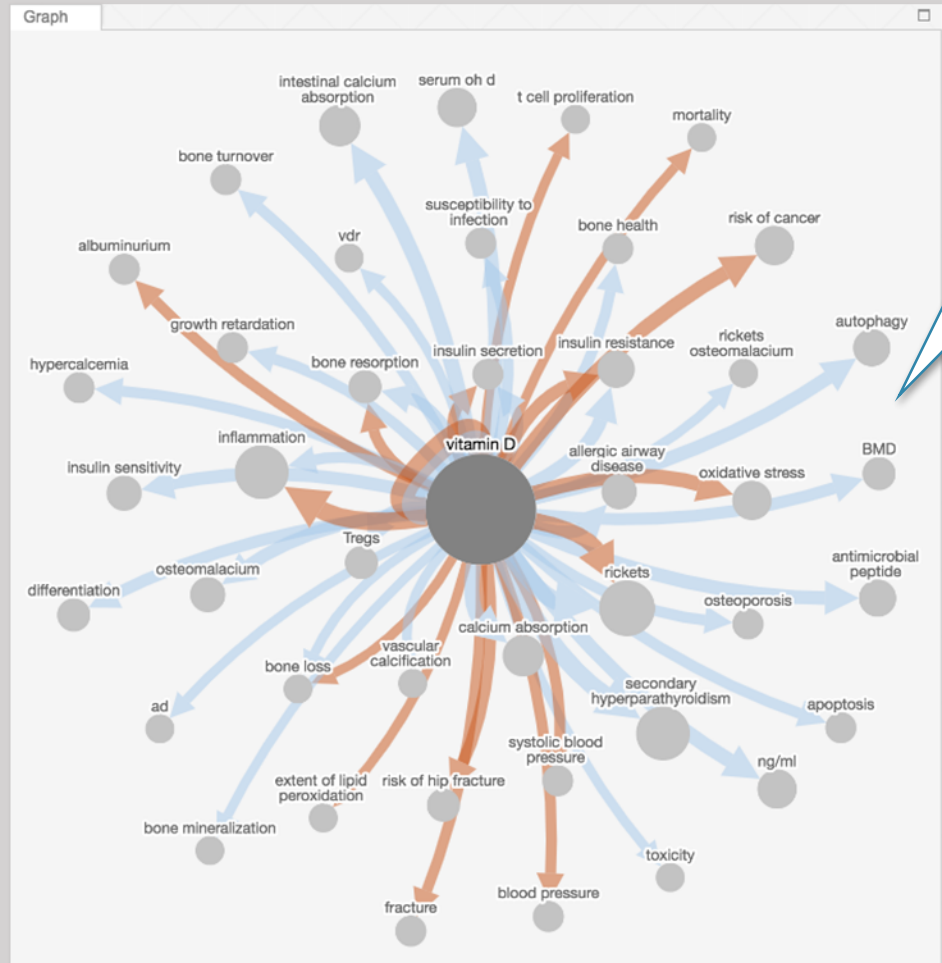
PMC4288271

. . . Iodine fortification of salt was followed in 1933 by the fortification of milk with vitamin D to prevent rickets based on the recommendation of the American Medical Association Council on Foods . . . .

PMC4377865

. . . Vitamin D prevents rickets by mediating calcium absorption in the intestines as well as calcium metabolism . . . .

Graph node labels: intestinal calcium absorption, serum calcium, bone turnover, vdr, susceptibility to infection, albuminuria, growth retardation, bone resorption, insulin secretion, insulin resistance, rickets, osteomalacium, hypercalcemia, inflammation, vitamin D, allergic airway disease, oxidative stress, BMD, insulin sensitivity, Tregs, antimicrobial peptide, differentiation, osteomalacium, rickets, osteoporosis, bone loss, vascular calcification, calcium absorption, secondary hyperparathyroidism, apoptosis, ad, extent of lipid peroxidation, risk of hip fracture, systolic blood pressure, ng/ml, bone mineralization, toxicity, fracture, blood pressure

Network graph view of the results, indicating influence relations between concepts. Edge and node thickness corresponds to the evidence count.
Orange edges indicate inhibition.
Blue edges indicate promotion.

Influence Search

Context | vitamin D | 1 | Effect | Search | ⚙ Options

Graph

Results

Toggle columns | Copy | CSV | Excel | Filter results:

Cause ⬍	Relation ⬍	Effect ⬍	Evidence ⬍
Filter...	Filter...	Filter...	Filter...
vitamin D	decreases	rickets	seen 12 times
vitamin D	increases	calcium absorption	seen 10 times
vitamin D	increases	intestinal calcium absorption	seen 10 times

Showing 1 to 50 of 3,089 entries   Previous 1 2 3 4 5 ... 62 Next

Show 50 ⏷ entries

Table of results. Each row corresponds to an edge in the graph. Results can be further filtered by applying searches to individual columns or globally. Data can be exported to other table formats.

Results can sorted by amount of evidence, or by the **likelihood of inter-disciplinary interest** (work in progress)

Evidence ✕

vitamin D *decreases* rickets

PMC4160567

. . . Adequate vitamin D is essential to prevent rickets , but adequate calcium is equally important ; if either calcium or vitamin D is deficient , bone health suffers . . . .
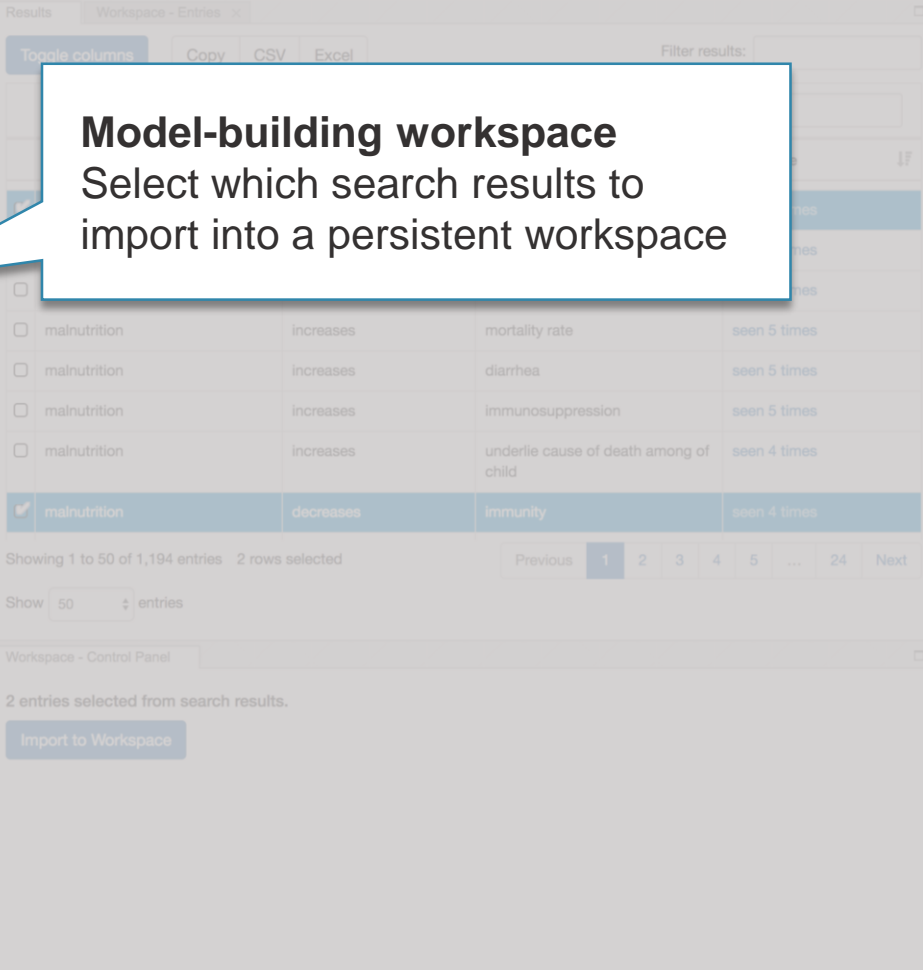
PMC4288271

. . . Iodine fortification of salt was followed in 1933 by the fortification of milk with vitamin D to prevent rickets based on the recommendation of the American Medical Association Council on Foods . . . .

PMC4377865

. . . Vitamin D prevents rickets by mediating calcium absorption in the intestines as well as calcium metabolism . . . .

Influence Search

Context | vitamin D | 1 | Effect | Search | ⚙ Options

Graph | Workspace - Graph ✕

Results | Workspace - Entries ✕

Toggle columns | Copy | CSV | Excel | Filter results:

Model-building workspace
Select which search results to import into a persistent workspace

	malnutrition	increases	mortality rate	seen 5 times
	malnutrition	increases	diarrhea	seen 5 times
	malnutrition	increases	immunosuppression	seen 5 times
	malnutrition	increases	underlie cause of death among of child	seen 4 times
✓	malnutrition	decreases	immunity	seen 4 times

Showing 1 to 50 of 1,194 entries   2 rows selected   Previous | 1 | 2 | 3 | 4 | 5 | ... | 24 | Next

Show | 50 | entries

Workspace - Control Panel

2 entries selected from search results.

Import to Workspace

Live demo: The Influence Graph Explorer

- What are some direct causes of **malnutrition**?
- What are some direct effects of **breastfeeding**?
- How are **viruses** indirectly linked to **obesity**?
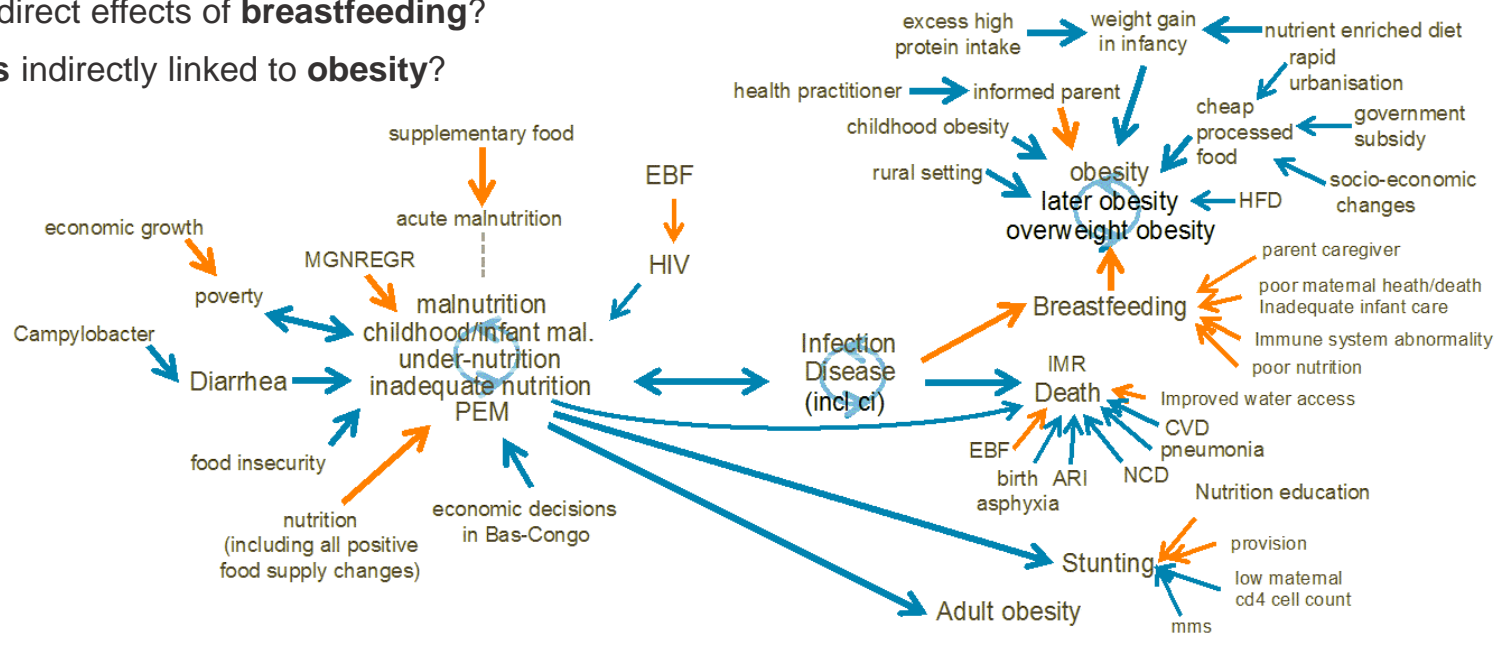
Model courtesy of: Lyn Powell.

HBGD*ki*

Live demo

Future Steps

- Ranking function that promotes multi-disciplinary discoveries
- Social network of science:
  - Collaboratively build multi-disciplinary models
- Beyond paywalls:
  - Work with publishers to integrate papers behind a paywall
- Integrate with other foundation tools
  - Kikm, Data Journey
- Beyond textual causality
  - Extract "associates with" relations

HBGD*ki*

Thank you.

Supplemental

HBGD*ki*

Bibliography

1. M.A. Valenzuela-Escarcega, G. Hahn-Powell, T. Hicks, and M. Surdeanu. *A Domain-independent Rule-based Framework for Event Extraction*. ACL-IJCNLP, 2015.

2. M. A. Valenzuela-Escarcega, G. Hahn-Powell, and M. Surdeanu. *Description of the Odin Event Extraction Framework and Rule Language*. arXiv:1509.07513, 2015.

3. D. Bell, G. Hahn-Powell, M. A. Valenzuela-Escarcega, and M. Surdeanu. *An Investigation of Coreference Phenomena in the Biomedical Domain*. LREC, 2016.

4. M. A. Valenzuela-Escarcega, G. Hahn-Powell, and M. Surdeanu. *Odin's Runes: A Rule Language for Information Extraction*. LREC, 2016.

5. G. Hahn-Powell, D. Bell, M. A. Valenzuela-Escarcega, and M. Surdeanu. *This before That: Influence Precedence in the Biomedical Domain*. BioNLP, 2016.

6. G. Hahn-Powell, M. A. Valenzuela-Escárcega, and M.Surdeanu. "Swanson linking revisited: Accelerating literature-based discovery across domains using a conceptual influence graph." *Proceedings of ACL 2017, System Demonstrations* (2017): 103-108.

7. M. A. Valenzuela-Escarcega, A. Nagesh, and M. Surdeanu. *Bootstrapped Representation Learning for Lexicon Acquisition*. In preparation.

8. M. A. Valenzuela-Escarcega, O. Babur, G. Hahn-Powell, D. Bell, T. Hicks, E. Noriega-Atala, X. Wang, M. Surdeanu, E. Demir, and C. Morrison. *Large-scale Automated Reading of Scientific Cancer Literature Discovers New Cancer Driving Mechanisms*. In preparation.