AI and ML in Healthcare

Lecture 01: An Overview

Vít Nováček

Faculty of Informatics, Masaryk University

Spring, 2023

MUNI

Outline

Outline

Scope of the Course

The photo's original source is the blog-post at `https://tinyurl.com/vxhbn5z4`, copyrighted by Asia Citro (license unknown, author unavailable for an explicit permission).

Your Background and Motivations

?

Structure of the Course (1/2)

- Weeks 01-02: overview of biomedical and healthcare informatics
  - Two introductory lectures:
    - Today is about relevant challenges and selected technologies
    - The next one will showcase possible solutions to selected problems
  - You'll also form working groups and get your first assignments
- Weeks 03-06: independent work in groups
  - An in-depth study of selected papers (one per group)
  - Coming up with a proposal of a project based on the studied paper
  - Weekly update meetings (online or in person) with each group
- Week 07: hackathon no. 1 (tentative topic: deep vs. classical ML)
  - In-person presentations of the results of the paper study
  - In-person presentations of the related project proposals
  - Kicking off the projects in the working groups
  - Possibly re-shuffling the groups if/as needed
  - Note: may last until late (bring refreshments!)

Structure of the Course (2/2)

- Weeks 08-09: independent work in groups
  - An in-depth study of selected papers (one per group, related to the previous one)
  - Progressing with the projects
  - Weekly update meetings (online or in person) with each group
- Week 10: hackathon no. 2 (tentative topic: structured or unstructured, that is the question)
  - In-person presentations of the results of the second paper study
  - An intense project session in the working groups
  - Note: once again, may last until late (bring refreshments!)
- Week 11: independent work in groups
  - Finalising the projects
  - Preparing the project presentations
  - Weekly update meetings (online or in person) with each group
- Week 12: in-person project presentations, colloquium
- Week 13: a closing session (hackathon no. 3, just refreshments, or even nothing at all, depending on the group consensus)

Assessment Criteria

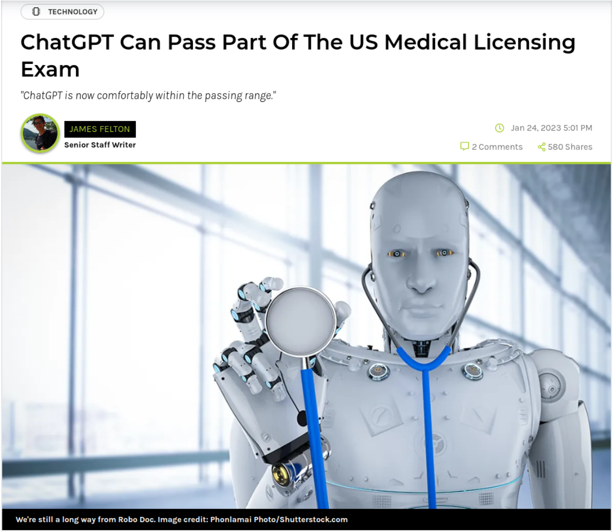
- Rather qualitative
- Based on active participation in
  - the study of the assigned papers
  - the weekly update meetings
  - the collaborative project work
  - the hackathons
- Corroborated by the final colloquium

Your Take on the Course Outline

?

Outline

Motivation?



TECHNOLOGY

ChatGPT Can Pass Part Of The US Medical Licensing Exam

"ChatGPT is now comfortably within the passing range."

JAMES FELTON
Senior Staff Writer

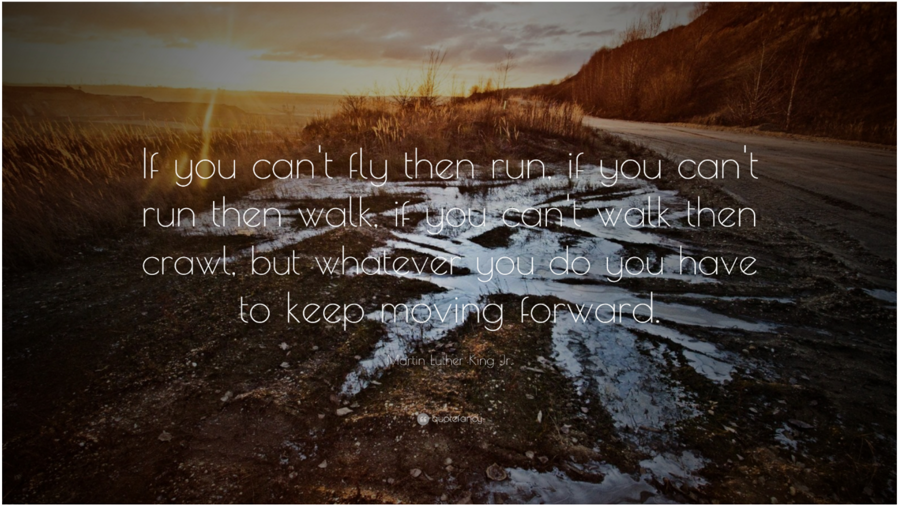
Jan 24, 2023 5:01 PM

2 Comments 580 Shares

We're still a long way from Robo Doc. Image credit: Phonlamai Photo/Shutterstock.com

Screenshot taken from an IFLScience article at `https://tinyurl.com/uau8yh84`.

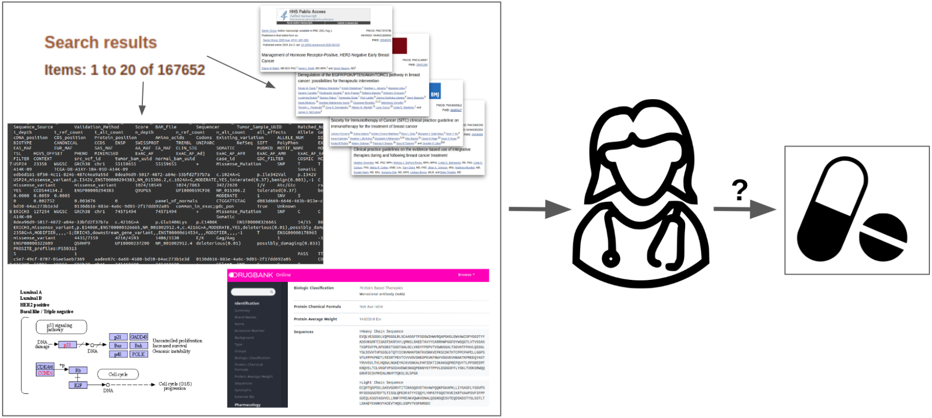
Motivation!



A Martin Luther King's quote, as framed in pictures at `https://tinyurl.com/ywavasyv` (license unknown).

Outline

The Perennial Grand Challenge – Information Overload



The icons' original sources: https://tinyurl.com/48b53fwc, https://tinyurl.com/3vhmx5xr (Public Domain). The part of a KEGG pathway taken from https://tinyurl.com/2x3c6bxz.

Outline

Genotype-Phenotype Mysteries

Malignant neoplasm of breast, C0006142

Source: ALL

1 - 25 of 6941 results

Results per page 25

Gene	UniProt	Gene Full Name	Protein Class	N. diseases$_{g}$	DSI$_{g}$	DPI$_{g}$	pLI	Score$_{gda}$	EL$_{gda}$
ESR1	P03372	estrogen receptor 1	Nuclear receptor	1101	0.324	0.962	1.00	1.000	None
BRCA1	P38398	BRCA1 DNA repair associa…	Enzyme	747	0.367	0.923	9.2E-29	1.000	strong
BRCA2	P51587	BRCA2 DNA repair associa…	Nucleic acid binding	656	0.379	0.846	2.4E-25	1.000	strong
TP53	P04637	tumor protein p53	Transcription factor	2494	0.236	0.962	0.53	1.000	None
PIK3CA	P42336	phosphatidylinositol-4,5-bis…	Kinase	1511	0.292	0.923	1.00	1.000	None

Results of a search using the DisGeNet database (c.f. `https://www.disgenet.org/`).

Protein Structure Puzzles

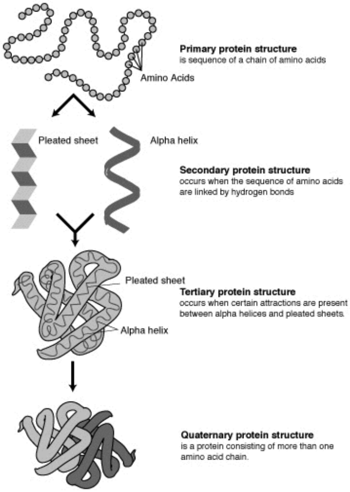


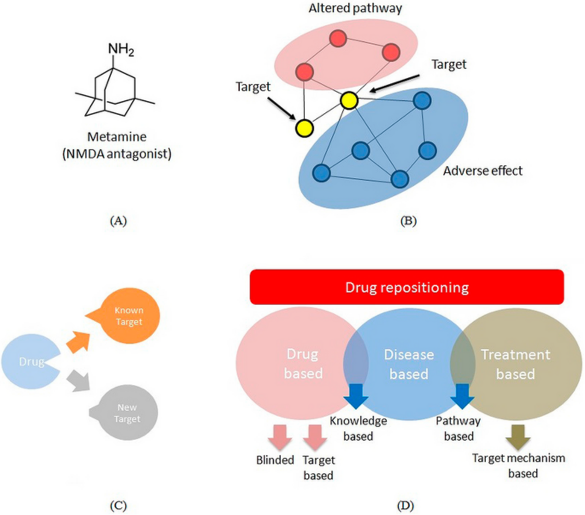
Image in Public Domain, courtesy of National Human Genome Research Institute.

Protein Function Conundrums

☐	AKT1S1	Proline-rich AKT1 substrate 1	protein binding	IntAct	Homo sapiens
☐	AKT1S1	Proline-rich AKT1 substrate 1	negative regulation of protein kinase activity	UniProt	Homo sapiens
☐	AKT1S1	Proline-rich AKT1 substrate 1	negative regulation of TOR signaling	UniProt	Homo sapiens
☐	AKT1S1	Proline-rich AKT1 substrate 1	regulation of apoptotic process	UniProt	Homo sapiens
☐	AKT1S1	Proline-rich AKT1 substrate 1	regulation of neuron apoptotic process	UniProt	Homo sapiens

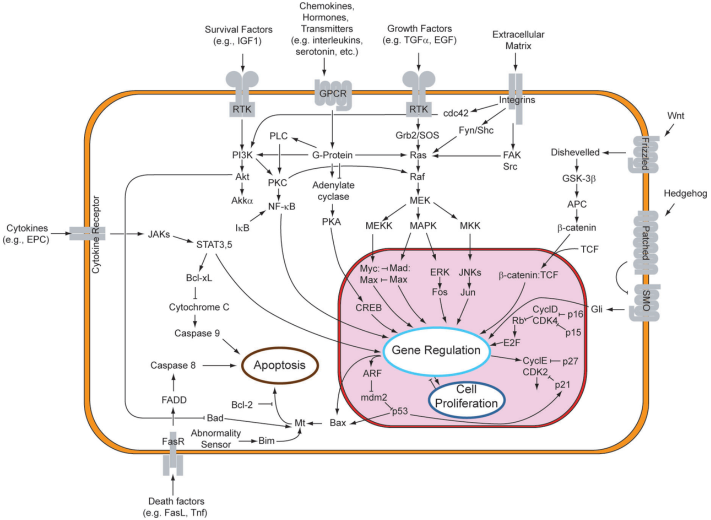
Results of a search using the Gene Ontology (c.f. `http://geneontology.org/`).

Needles and Haystacks of Drug Design



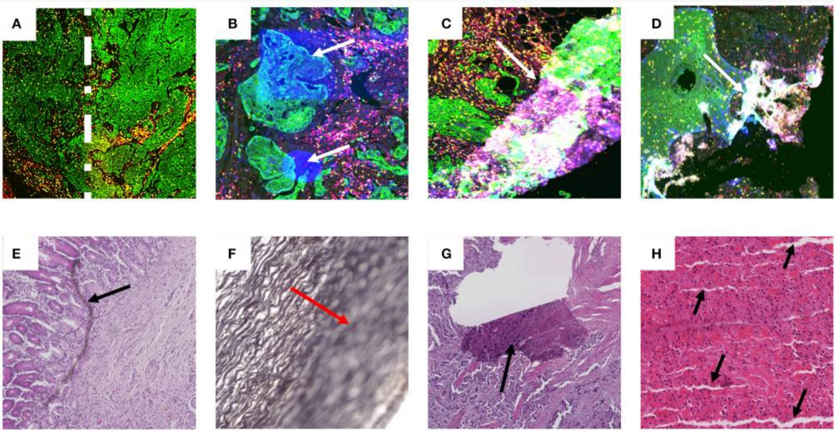
The drug target graphics comes from the special issue of the Biomedicines journal on Molecular Imaging as a Tool for Personalized Medicine (c.f. `https://tinyurl.com/yd9b6ryf` for details on attribution).

Signalling Tangles



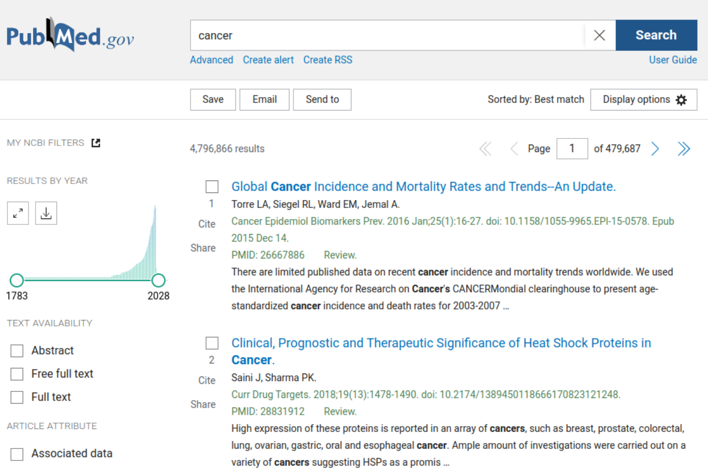
The image is based on `en:Image:Signal_transduction_pathways.jpg` from Wikimedia Commons which was released to the public domain by `en:User:Roadnottaken`.

Image Mazes

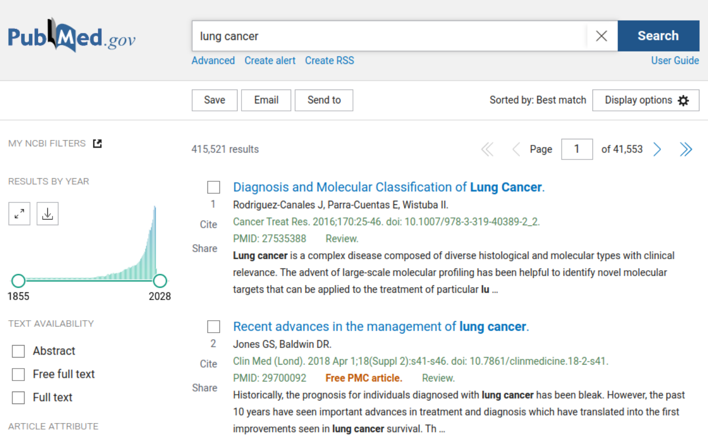


The WSI image originally provided in `https://doi.org/10.3389%2Ffmed.2019.00264`.

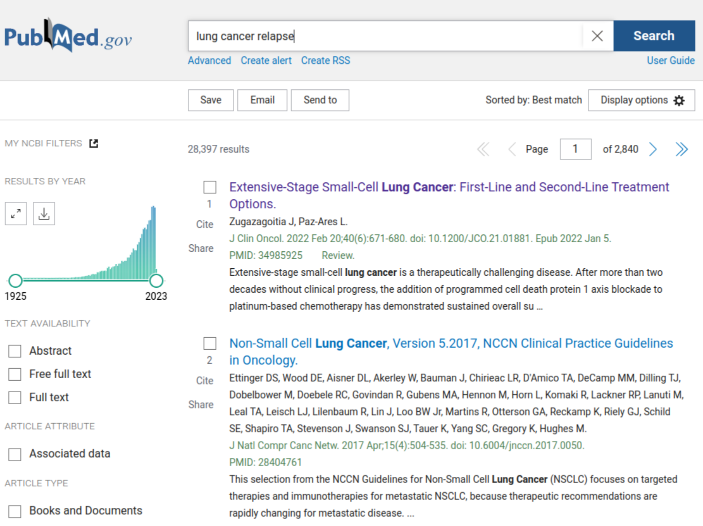
Publication Deluge (1/3)



Publication Deluge (2/3)



Publication Deluge (3/3)



Outline

A Bit of History – Expert Systems (1/2)

- A system for emulating human decision capabilities
- Two key parts
  - A knowledge base – facts and rules (typically IF-THEN in the US tradition, or more expressive First Order Logic in the European tradition)
  - An inference engine – an automated reasoning module for answering queries and/or deducing new information (typically backward/forward chaining or a Prolog interpreter)
- Foundational research in 1940s-1950s, official definition in 1960s, flourishing till 1980s
- Succumbed to the AI winter of the 1990s, resurrection in 2000s (rule-based systems developed by SAP, Oracle, etc.)
- Now more or less replaced by or integrated with data mining and (deep) machine learning frameworks

A Bit of History – Expert Systems (2/2)

- Medicine has been one of the most important driving factors
- MYCIN – the first big success story
  - Developed in 1970s by some of the leading figures in biomedical informatics
  - An expert system for identifying bacteria causing severe infections
  - KB of about 600 rules, backward chaining inference engine with uncertainty support
- Iterative QA process via series of simple yes/no questions to the physician
- Never actually used in practice
- Still, its treatment plans had high acceptability rating of 65% (comparable to a panel of 5 Stanford experts)

Outline

Biomedical Knowledge Representation – Why Bother?

Original source: `https://xkcd.com/1838`.

Controlled Vocabularies

- Resources for organising knowledge for subsequent retrieval
- Motivated by the need for standardised tagging of information units
  - Evolved in library and information science (paper-based at first)
  - Useful for machine-aided information retrieval, too
- Prerequisite for more sophisticated knowledge organisation systems
  - Subject headings, thesauri, taxonomies, ...
  - Ontologies, rule bases, knowledge graphs, ...

Ontologies

- Representation, formal naming and definition of general categories (also called concepts or classes) and individuals falling under them
- Properties of the categories and individual entities, relationships between them
- Metadata and annotations that do not affect the formal meaning
- Typically based on subsets of first order predicate logic, such as Description Logics
- May allow for expressive deductive reasoning
- Sophisticated, but pretty heavy-weight and expensive to create and maintain

Interoperability Standards (1/2)

- Specifications for representation, storing and exchange of data, information and knowledge
- Defined and agreed on by relevant stakeholders (both from academia and industry)
- Accepted and used by a broad community of independent practicians

Interoperability Standards (2/2)

- Examples of standards endorsed by US-based bodies:
  - *LOINC* – health measurements, observations, and documents; *RxNorm* – normalized names for clinical drugs; *SNOMED CT* – set of standards for electronic exchange of clinical health information; *UMLS* – integrates key terminology, classification and coding standards (all NLM)
  - *Standard for Safety for Medical Device Interoperability* (FDA/ANSI)
  - *Digital Imaging and Communications in Medicine* (DICOM; National Electrical Manufacturers Association)
  - *HL7* – a set of international standards for transfer of clinical and administrative data between software applications (HL7 International/ANSI)
- Examples of standards endorsed by EU-based bodies:
  - *ICD-11* – global standard for recording health information and causes of death (WHO)
  - *MedDRA* – clinically validated international medical terminology dictionary-thesaurus, primarily used in pharmacology, pharmacovigilance and clinical research (ICH)

Biomedical KR Example – MeSH

- Temporal Central Focal Epilepsy
- Temporal-Central Focal Epilepsies
- Benign Childhood Epilepsy With Centro-Temporal Spikes
- Benign Childhood Epilepsy With Centro Temporal Spikes
- BCECTS

Previous Indexing:

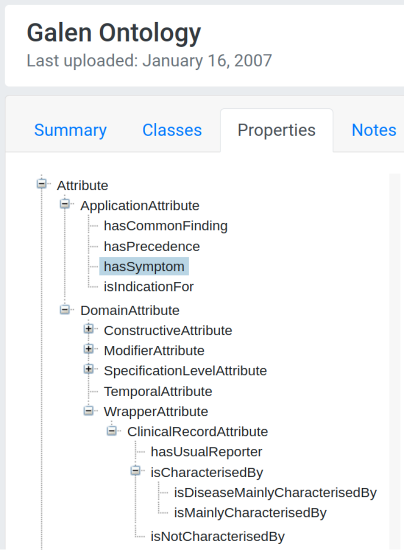
- Epilepsy, Partial (1986-1996)

All MeSH Categories
  Diseases Category
    Nervous System Diseases
      Central Nervous System Diseases
        Brain Diseases
          Epilepsy
            Epilepsies, Partial
              Epilepsy, Rolandic

All MeSH Categories
  Diseases Category
    Nervous System Diseases
      Central Nervous System Diseases
        Brain Diseases
          Epilepsy
            Epileptic Syndromes
              Epilepsy, Rolandic

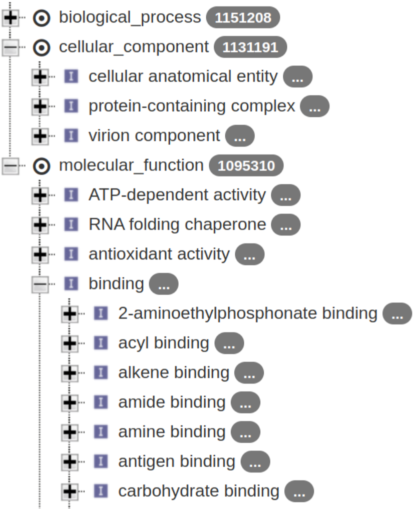
Screenshot taken from the `https://tinyurl.com/yckt72zs` query results.

Biomedical KR Example – GALEN



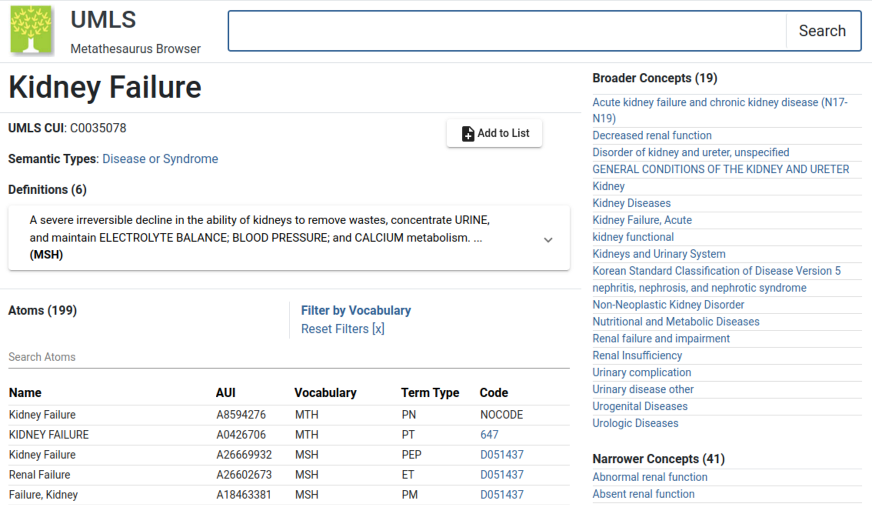
Screenshot taken from NCBO's BioPortal (c.f., `https://tinyurl.com/ye34eb4p`).

Biomedical KR Example – Gene Ontology



Screenshot taken from the AmiGO browser (c.f., `http://amigo.geneontology.org/`).

Biomedical KR Example – UMLS



UMLS
Metathesaurus Browser

Search

Kidney Failure

UMLS CUI: C0035078

Add to List

Semantic Types: Disease or Syndrome

Definitions (6)

A severe irreversible decline in the ability of kidneys to remove wastes, concentrate URINE, and maintain ELECTROLYTE BALANCE; BLOOD PRESSURE; and CALCIUM metabolism. ... **(MSH)**

Atoms (199)

Filter by Vocabulary
Reset Filters [x]

Search Atoms

Name	AUI	Vocabulary	Term Type	Code
Kidney Failure	A8594276	MTH	PN	NOCODE
KIDNEY FAILURE	A0426706	MTH	PT	647
Kidney Failure	A26669932	MSH	PEP	D051437
Renal Failure	A26602673	MSH	ET	D051437
Failure, Kidney	A18463381	MSH	PM	D051437

Broader Concepts (19)

- Acute kidney failure and chronic kidney disease (N17-N19)
- Decreased renal function
- Disorder of kidney and ureter, unspecified
- GENERAL CONDITIONS OF THE KIDNEY AND URETER
- Kidney
- Kidney Diseases
- Kidney Failure, Acute
- kidney functional
- Kidneys and Urinary System
- Korean Standard Classification of Disease Version 5
- nephritis, nephrosis, and nephrotic syndrome
- Non-Neoplastic Kidney Disorder
- Nutritional and Metabolic Diseases
- Renal failure and impairment
- Renal Insufficiency
- Urinary complication
- Urinary disease other
- Urogenital Diseases
- Urologic Diseases

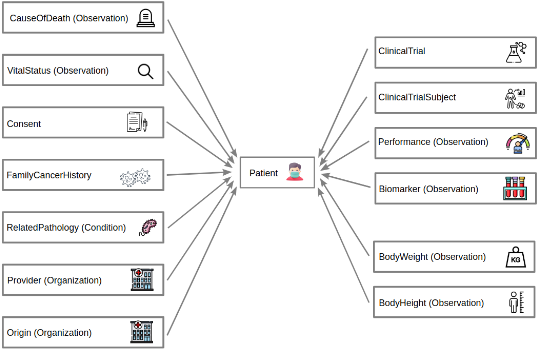
Narrower Concepts (41)

- Abnormal renal function
- Absent renal function

Screenshot taken from `https://tinyurl.com/w2pf3kra`.

Biomedical KR Example – OSIRIS (HL7/FHIR)

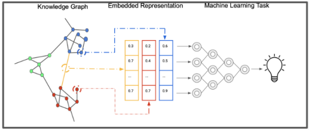
- Interoperability and data sharing framework for clinical and biological data in oncology
- Established in France in 2015, re-used by a number of EU stakeholders since then



Screenshot of the top-level OSIRIS schema taken from `https://fhir.arkhn.com/osiris/`.

Knowledge Graphs as an ML-Ready KR Paradigm

$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$
$\bot^{\mathcal{I}} = \emptyset$
$(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$ *(union means disjunction)*
$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$ *(intersection means conjunction)*
$(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$ *(complement means negation)*
$(\forall R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} | \text{for every } y, (x,y) \in R^{\mathcal{I}} \text{ implies } y \in C^{\mathcal{I}}\}$
$(\exists R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} | \text{there exists } y, (x,y) \in R^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\}$



- Somewhat related to ontologies, but departing from formal semantics
  - A lightweight, yet powerful graph-based KR, organising descriptions of objects and their connections/properties
  - Data instead of schema, bottom-up instead of top-down, transduction instead of deduction, . . .
- Applicable to many domains and use cases
- Straightforward automated population and knowledge integration
- Rather complex inference still possible
  - Link prediction, knowledge base completion, relation extraction, analogical reasoning
  - Schema induction, FOL / DL axioms can be incorporated
- Scalable algorithms taking advantage of the deep learning state of the art

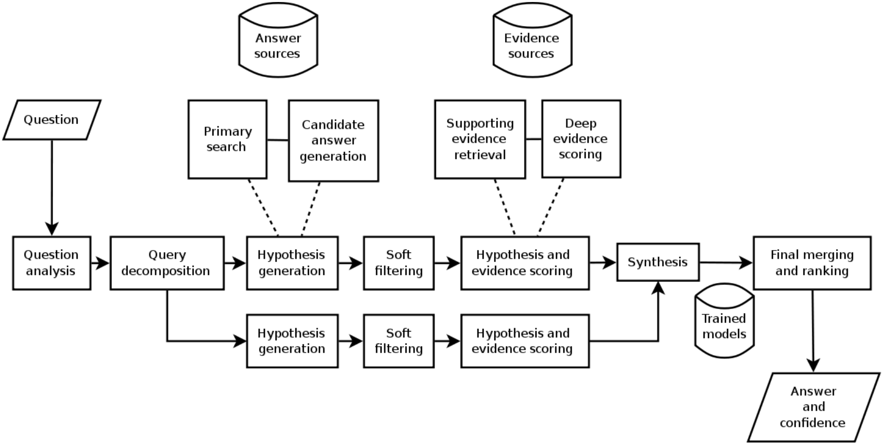
The DL semantics list's original source: https://tinyurl.com/53khkptu. The KG embedding graphics created by Edoardo Ramalli - Own work, CC BY-SA 4.0, https://tinyurl.com/3dr7c76b.

Outline

IBM Watson for Oncology – How It Started

Original image source: a copyrighted TV show screenshot (a non-free content, used according to the fair use policy under US copyright law; for more info, c.f. `https://tinyurl.com/yvpjspyr`).

IBM Watson for Oncology – How It Worked

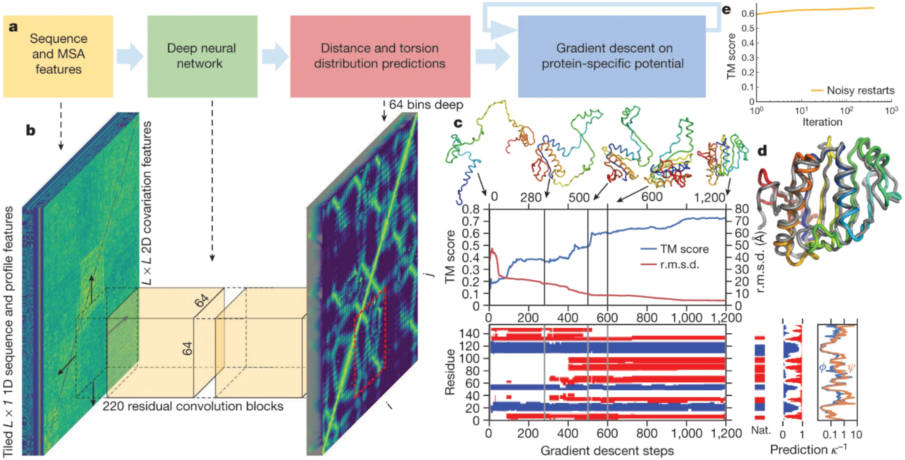


Original image source: `https://tinyurl.com/bjuu3chm` (own work by the Pgr94 Wikipedia user, released under CC0 1.0 Universal Public Domain Dedication).

IBM Watson for Oncology – How It's Going

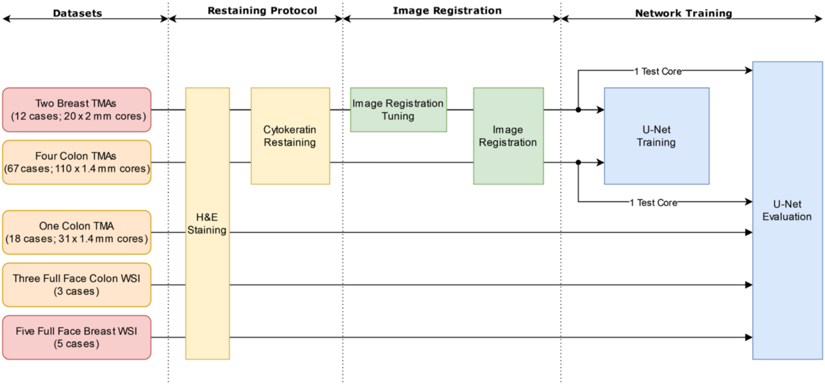
Screenshots taken from `https://tinyurl.com/3pt4j32y`, `https://tinyurl.com/yckvkcc7`, `https://tinyurl.com/2eky7nvk`, `https://tinyurl.com/2kvbmc2h`.

AlphaFold – A Breakthrough in Protein Structure Prediction



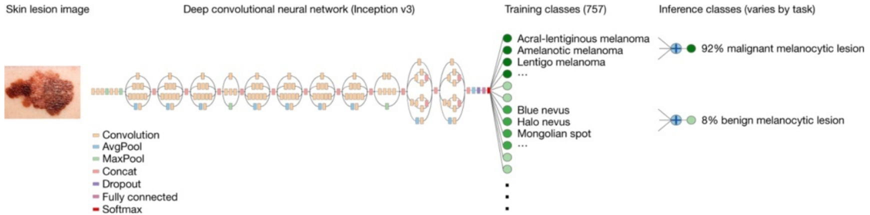
Original image source: the corresponding research paper at https://doi.org/10.1038/s41586-019-1923-7.

Digital Pathology – Automated Annotations of Epithelial Cells and Stroma



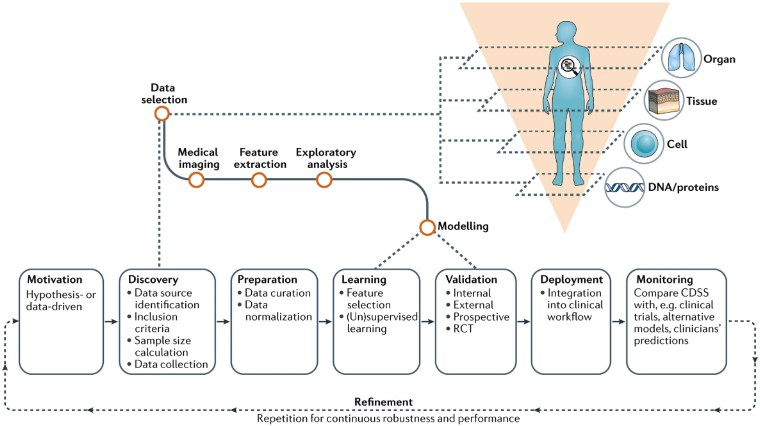
Original image source: the corresponding research paper at https://doi.org/10.1002/cjp2.249.

Deep Learning in Melanoma Diagnostics



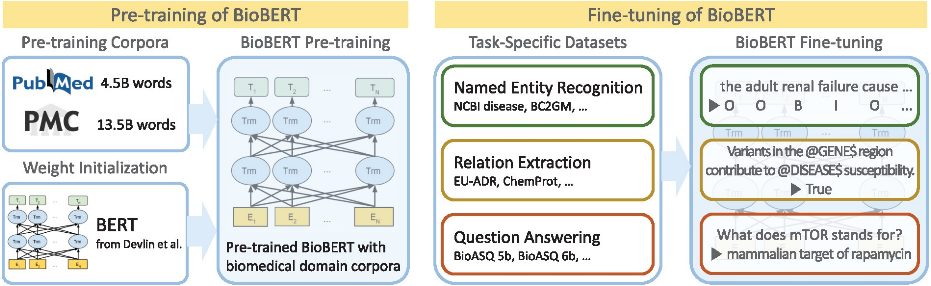
Original image source: the corresponding research paper at `https://doi.org/10.1038/nature21056`.

Radiomics – Clinical Decision Support Based on Imaging Data



Original image source: the corresponding research paper at `https://doi.org/10.1038/nrclinonc.2017.141`.

Biomedical Language Models – Making Use of the State of the Art in “AI”



Original image source: the corresponding research paper at `https://doi.org/10.1093/bioinformatics/btz682`.

Outline

Related Overall Readings on the Field and Biomedical KR

- Shortliffe, Edward H., et al. Biomedical informatics: computer applications in health care and biomedicine. Springer, 2014.
- Buchanan, Bruce G., and Edward H. Shortliffe. Rule based expert systems: the Mycin experiments of the Stanford Heuristic programming project (the Addison-Wesley series in artificial intelligence). Addison-Wesley Longman Publishing Co., Inc., 1984.
- Smith, Barry, et al. "Relations in biomedical ontologies." Genome biology 6.5 (2005): 1-15.
- Musen, Mark A., et al. "The national center for biomedical ontology." Journal of the American Medical Informatics Association 19.2 (2012): 190-195.
- Hoehndorf, Robert, Michel Dumontier, and Georgios V. Gkoutos. "Evaluation of research in biomedical ontologies." Briefings in bioinformatics 14.6 (2013): 696-712.
- Li, Michelle M., Kexin Huang, and Marinka Zitnik. "Graph representation learning in biomedicine and healthcare." Nature Biomedical Engineering (2022): 1-17.

Related Readings on (Deep) Machine Learning

- Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. "Machine learning in medicine." New England Journal of Medicine 380.14 (2019): 1347-1358.
- Ching, Travers, et al. "Opportunities and obstacles for deep learning in biology and medicine." Journal of The Royal Society Interface 15.141 (2018): 20170387.
- Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." nature 542.7639 (2017): 115-118.
- Senior, Andrew W., et al. "Improved protein structure prediction using potentials from deep learning." Nature 577.7792 (2020): 706-710.
- Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." Nature 596.7873 (2021): 583-589.
- Lambin, Philippe, et al. "Radiomics: the bridge between medical imaging and personalized medicine." Nature reviews Clinical oncology 14.12 (2017): 749-762.
- Brázdil, Tomáš, et al. "Automated annotations of epithelial cells and stroma in hematoxylin–eosin-stained whole-slide images using cytokeratin re-staining." The Journal of Pathology: Clinical Research 8.2 (2022): 129-142.

Related Readings on Language Technologies

- Cohen, Aaron M., and William R. Hersh. "A survey of current work in biomedical text mining." Briefings in bioinformatics 6.1 (2005): 57-71.
- Neumann, Mark, et al. "ScispaCy: fast and robust models for biomedical natural language processing." arXiv preprint arXiv:1902.07669 (2019).
- Lee, Jinhyuk, et al. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." Bioinformatics 36.4 (2020): 1234-1240.
- Gu, Yu, et al. "Domain-specific language model pretraining for biomedical natural language processing." ACM Transactions on Computing for Healthcare (HEALTH) 3.1 (2021): 1-23.

Related Readings on Clinical Decision Support Systems

- Berner, Eta S. Clinical decision support systems. Vol. 233. New York: Springer Science+ Business Media, LLC, 2007.
- Bright, Tiffani J., et al. "Effect of clinical decision-support systems: a systematic review." Annals of internal medicine 157.1 (2012): 29-43.
- Yang, Qian, Aaron Steinfeld, and John Zimmerman. "Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes." Proceedings of the 2019 CHI conference on human factors in computing systems. 2019.
- Antoniadi, Anna Markella, et al. "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review." Applied Sciences 11.11 (2021): 5088.
- Hanajíková, Michaela. "A review of clinical decision support systems applicable to precision oncology." BS thesis, Faculty of Informatics, Masaryk University (2022). Text available at: `https://is.muni.cz/auth/th/o8vzq/hanajikova_michaela_thesis_final.pdf`

Outline

Summary and Outlook

- Covered today:
  - Overview of the course
  - Overview of the field
  - Selection of further readings
- Outline of the next week:
  - Using AI/ML technologies – selected examples
  - Finalising the course structure, meeting times, etc.
  - Finalising the working groups
  - Assignment of papers to the groups

Action Items

- Everyone:
  - Put together a list of individual interests relevant to the field (now)
  - Identify our "clinical experts" central to each working group (now)
- I:
  - Send a poll on class re-scheduling preferences (by Thu 16/02)
  - Send a poll on progress meeting preferences (by Thu 16/02)
  - Suggest a viable re-scheduling of the classes (by Sun 19/02)
  - Compile the first batch of papers to study (by Wed 22/02)
- You:
  - Give me your class and meeting time preferences (by Fri 17/02)
  - Form working groups (by Wed 22/02)

Questions

?