Group Project

Satya Almasian/Ashish Chouhan
Data Science Group
Heidelberg University
October 19, 2023

Overview

- Composition of Course Grade
- Bi-weekly assignments
- Groups
- Project Topic and Data
- Project Timeline
- Repository and Report
- Mentors
- Final Notes

Composition of Course Grade

- Bi-weekly assignments (35% of the final grade); about 4 assignments

- Group project (65% of the final grade)

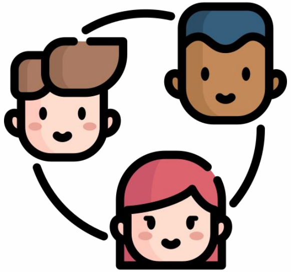
“Bi-weekly” Assignments

- Improve theoretical and practical skills
- Written answers (text, calculation, …) and programming
- All assignments are due Mondays (2pm before the lecture) via moodle
- First assignment will be out on 30.10.2023
- One submission per group (**3 to 4 students**)
- Write the name of all team members on the sheet.

Group Project

- Milestone description (10%)
- Project report + evaluation (40%)
- Code + running system (50%)

Groups

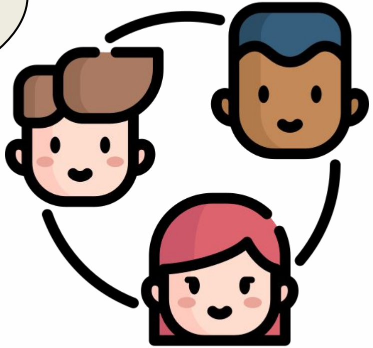


- Create group after this tutorial or by using moodle.
- Same group for assignment and project
  (form groups until 23.10.2023)
- Everyone should contribute!
- Grades are assigned per person not per group.

Project Topic

Imagine you are a group of NLP specialists, and we are a company in need of a **Question Answering** (QA) system:

Use the dataset provided to create a system to answer questions about our data in natural language.



Data

Medical Data:

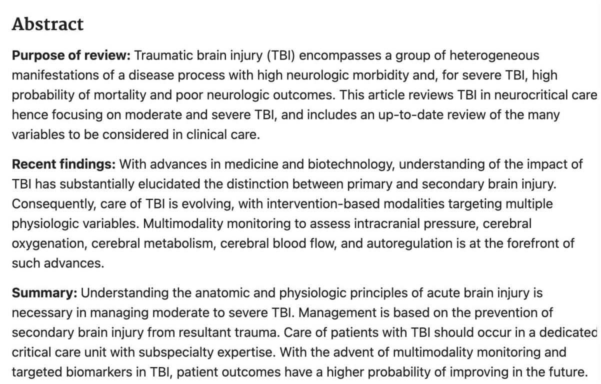
Selection of PubMed Articles [Link]

- Abstracts concerning “intelligence” from 2013 till 2023

Legal Data:

Selection of Eur-lex Articles [Link]

- Articles concerning “energy” from 1949 till 2023

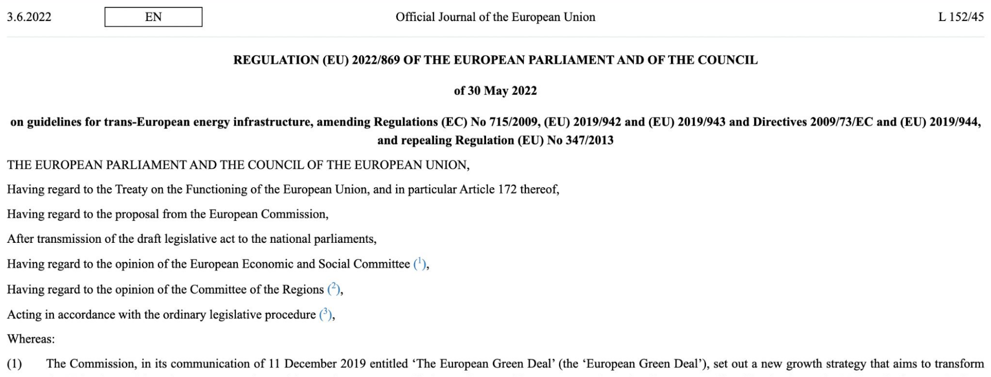


Abstract

Purpose of review: Traumatic brain injury (TBI) encompasses a group of heterogeneous manifestations of a disease process with high neurologic morbidity and, for severe TBI, high probability of mortality and poor neurologic outcomes. This article reviews TBI in neurocritical care hence focusing on moderate and severe TBI, and includes an up-to-date review of the many variables to be considered in clinical care.

Recent findings: With advances in medicine and biotechnology, understanding of the impact of TBI has substantially elucidated the distinction between primary and secondary brain injury. Consequently, care of TBI is evolving, with intervention-based modalities targeting multiple physiologic variables. Multimodality monitoring to assess intracranial pressure, cerebral oxygenation, cerebral metabolism, cerebral blood flow, and autoregulation is at the forefront of such advances.

Summary: Understanding the anatomic and physiologic principles of acute brain injury is necessary in managing moderate to severe TBI. Management is based on the prevention of secondary brain injury from resultant trauma. Care of patients with TBI should occur in a dedicated critical care unit with subspecialty expertise. With the advent of multimodality monitoring and targeted biomarkers in TBI, patient outcomes have a higher probability of improving in the future.



3.6.2022 EN Official Journal of the European Union L 152/45

REGULATION (EU) 2022/869 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 30 May 2022

on guidelines for trans-European energy infrastructure, amending Regulations (EC) No 715/2009, (EU) 2019/942 and (EU) 2019/943 and Directives 2009/73/EC and (EU) 2019/944, and repealing Regulation (EU) No 347/2013

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 172 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee ([1]),

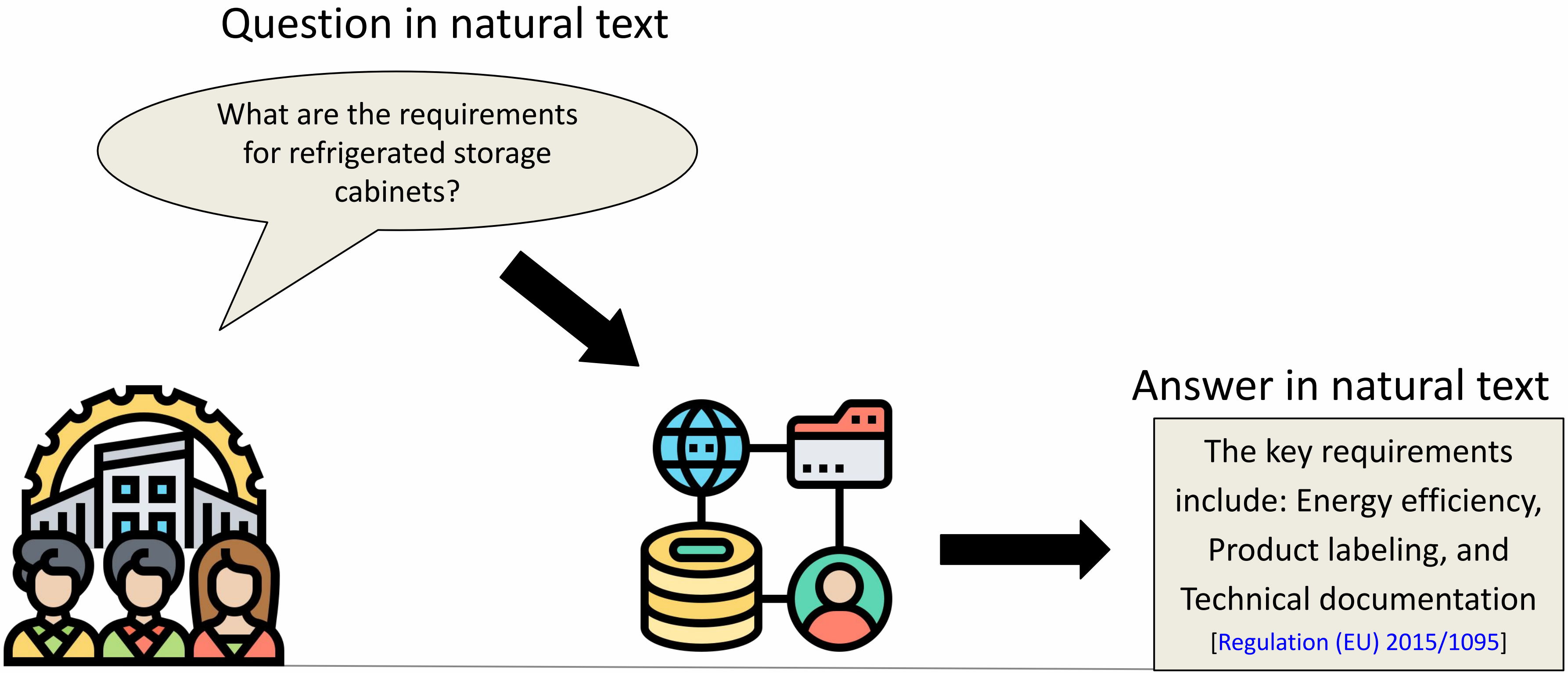
Having regard to the opinion of the Committee of the Regions ([2]),

Acting in accordance with the ordinary legislative procedure ([3]),

Whereas:

(1) The Commission, in its communication of 11 December 2019 entitled ‘The European Green Deal’ (the ‘European Green Deal’), set out a new growth strategy that aims to transform

Input and Output



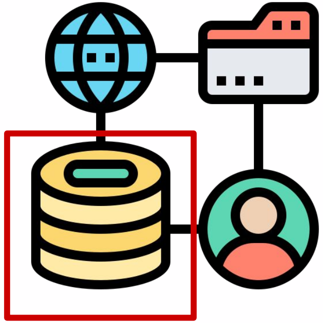
Question in natural text

What are the requirements for refrigerated storage cabinets?

Answer in natural text

The key requirements include: Energy efficiency, Product labeling, and Technical documentation [Regulation (EU) 2015/1095]

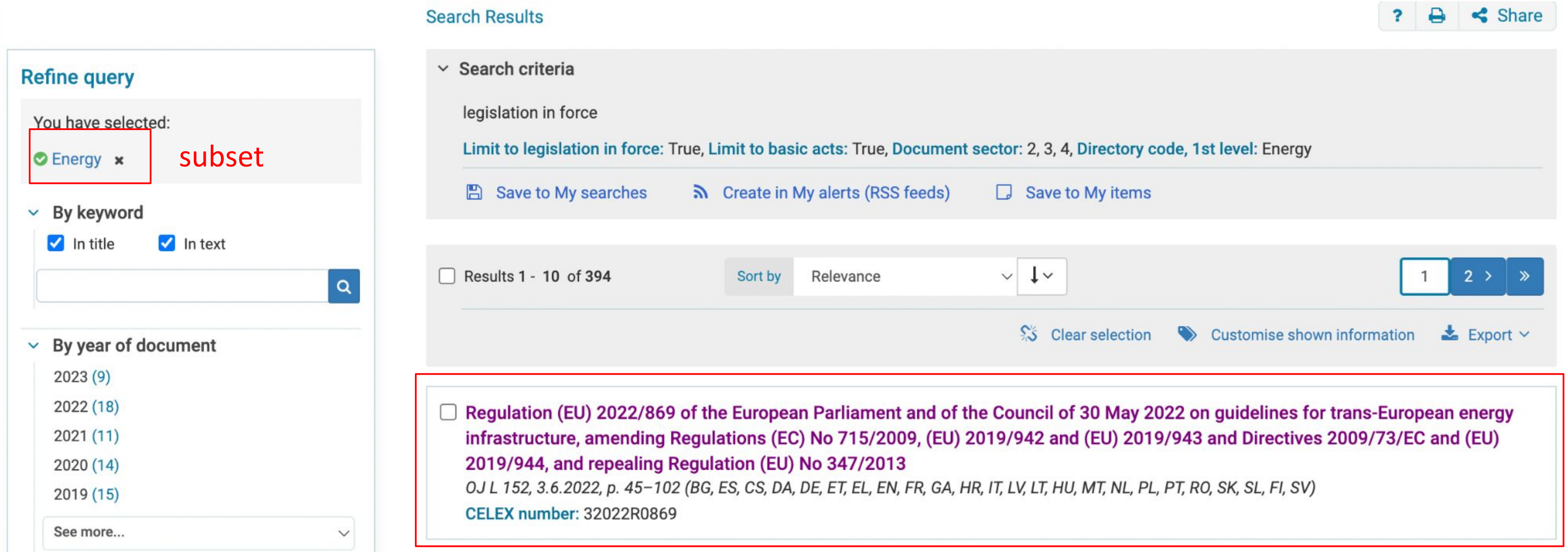
How to Build such System? (Step 1)



- Start with data acquisition.
- Collect the document subset needed.
- Find a way to properly store the data.
- Find a way to search through the documents.

Eur-lex Example

Eur-Lex: official website of European Union law, provides access to EU legal documents in 24 languages



Search Results

Refine query

You have selected:

Energy ×

subset

By keyword

☑ In title ☑ In text

By year of document

2023 (9)

2022 (18)

2021 (11)

2020 (14)

2019 (15)

See more...

Search criteria

legislation in force

Limit to legislation in force: True, Limit to basic acts: True, Document sector: 2, 3, 4, Directory code, 1st level: Energy

□ Results 1 - 10 of 394

Sort by: Relevance

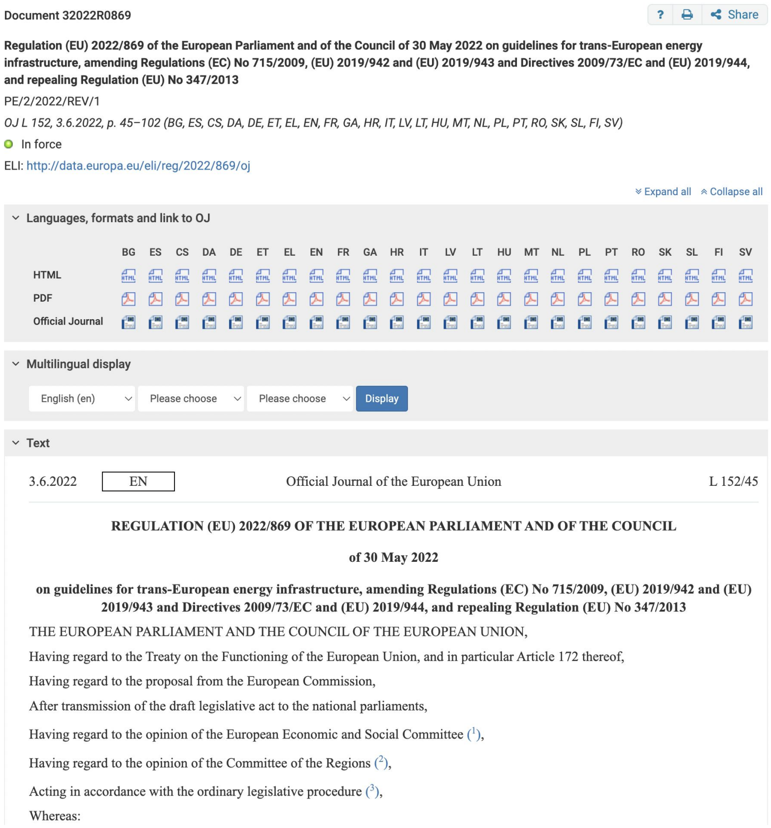
□ Regulation (EU) 2022/869 of the European Parliament and of the Council of 30 May 2022 on guidelines for trans-European energy infrastructure, amending Regulations (EC) No 715/2009, (EU) 2019/942 and (EU) 2019/943 and Directives 2009/73/EC and (EU) 2019/944, and repealing Regulation (EU) No 347/2013

OJ L 152, 3.6.2022, p. 45–102 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)

CELEX number: 32022R0869

Documents in multiple languages

Eur-lex Example



Document 32022R0869

Regulation (EU) 2022/869 of the European Parliament and of the Council of 30 May 2022 on guidelines for trans-European energy infrastructure, amending Regulations (EC) No 715/2009, (EU) 2019/942 and (EU) 2019/943 and Directives 2009/73/EC and (EU) 2019/944, and repealing Regulation (EU) No 347/2013

PE/2/2022/REV/1

OJ L 152, 3.6.2022, p. 45–102 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)

● In force

ELI: http://data.europa.eu/eli/reg/2022/869/oj

Languages, formats and link to OJ

	BG	ES	CS	DA	DE	ET	EL	EN	FR	GA	HR	IT	LV	LT	HU	MT	NL	PL	PT	RO	SK	SL	FI	SV
HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML	HTML
PDF																								
Official Journal																								

Multilingual display

Text

3.6.2022 EN Official Journal of the European Union L 152/45

REGULATION (EU) 2022/869 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 30 May 2022

on guidelines for trans-European energy infrastructure, amending Regulations (EC) No 715/2009, (EU) 2019/942 and (EU) 2019/943 and Directives 2009/73/EC and (EU) 2019/944, and repealing Regulation (EU) No 347/2013

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 172 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee [1],

Having regard to the opinion of the Committee of the Regions [2],

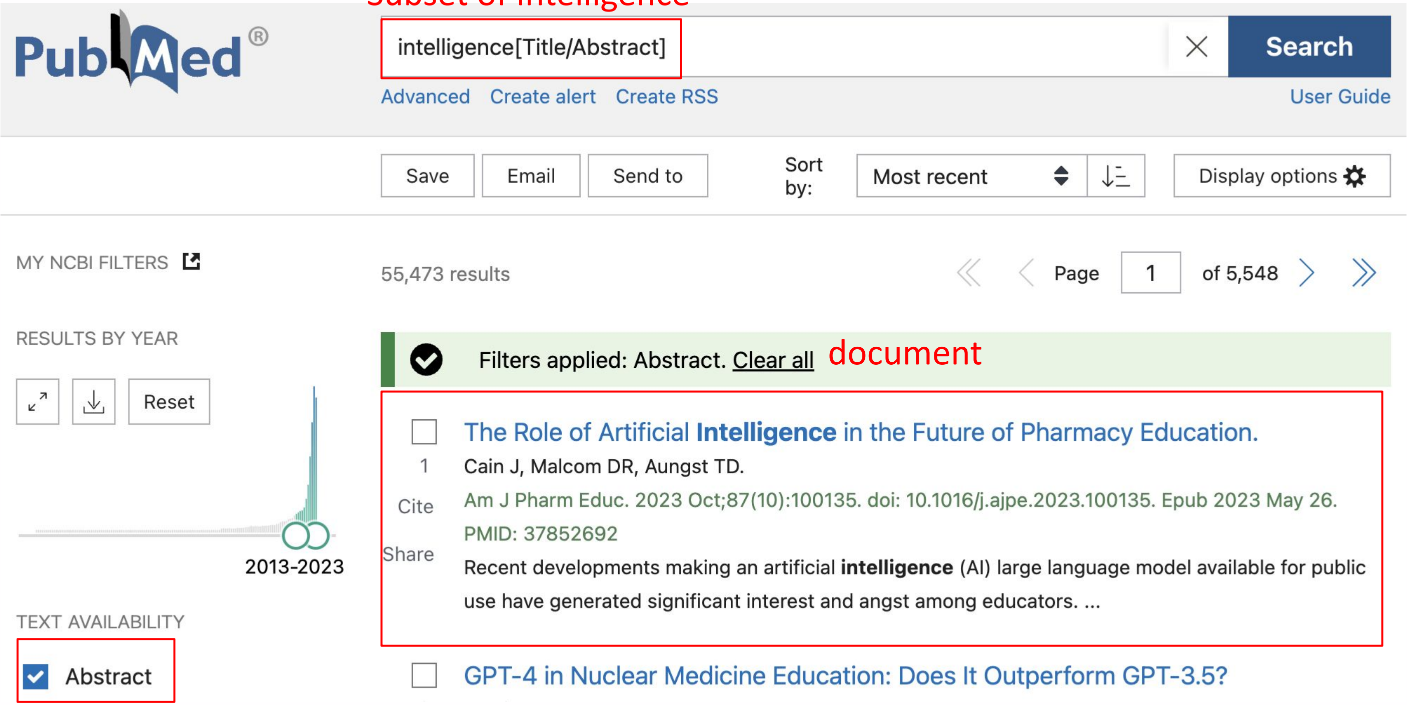
Acting in accordance with the ordinary legislative procedure [3],

Whereas:

- What metadata is needed?
- Only english language
- Structure of the document matters
  - Entire text
  - Paragraph
  - Sentences
- Choose the database according to the model used in later stages.

PubMed Example

PubMed: contains citations for biomedical literature from MEDLINE, life science journals, and online books.



Subset of intelligence

document

Only abstracts

intelligence[Title/Abstract]

55,473 results

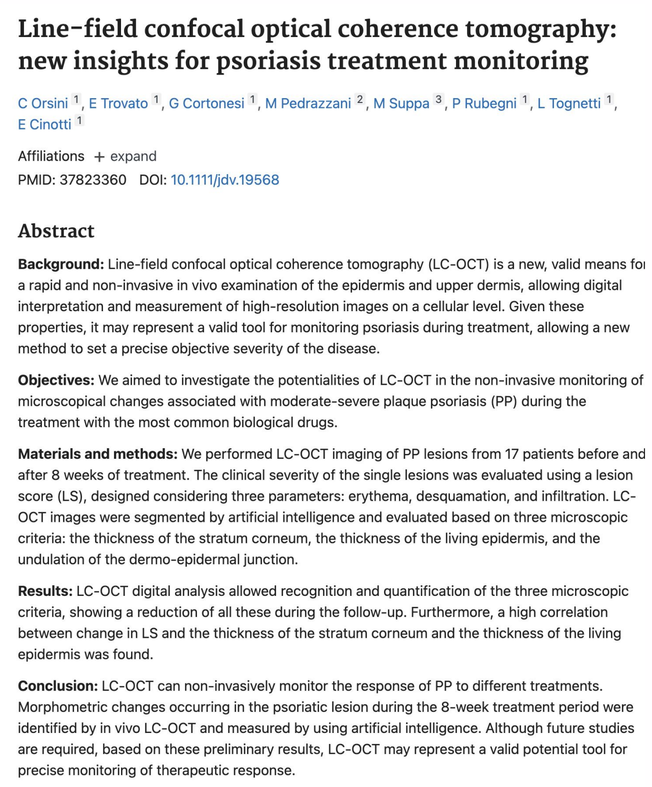
☑ Abstract

Filters applied: Abstract. Clear all

□ 1 The Role of Artificial **Intelligence** in the Future of Pharmacy Education.
Cain J, Malcom DR, Aungst TD.
Am J Pharm Educ. 2023 Oct;87(10):100135. doi: 10.1016/j.ajpe.2023.100135. Epub 2023 May 26.
PMID: 37852692
Recent developments making an artificial **intelligence** (AI) large language model available for public use have generated significant interest and angst among educators. ...

□ GPT-4 in Nuclear Medicine Education: Does It Outperform GPT-3.5?

PubMed Example



Line-field confocal optical coherence tomography: new insights for psoriasis treatment monitoring

C Orsini [1], E Trovato [1], G Cortonesi [1], M Pedrazzani [2], M Suppa [3], P Rubegni [1], L Tognetti [1], E Cinotti [1]

Affiliations + expand

PMID: 37823360 DOI: 10.1111/jdv.19568

Abstract

Background: Line-field confocal optical coherence tomography (LC-OCT) is a new, valid means fo a rapid and non-invasive in vivo examination of the epidermis and upper dermis, allowing digital interpretation and measurement of high-resolution images on a cellular level. Given these properties, it may represent a valid tool for monitoring psoriasis during treatment, allowing a new method to set a precise objective severity of the disease.

Objectives: We aimed to investigate the potentialities of LC-OCT in the non-invasive monitoring of microscopical changes associated with moderate-severe plaque psoriasis (PP) during the treatment with the most common biological drugs.

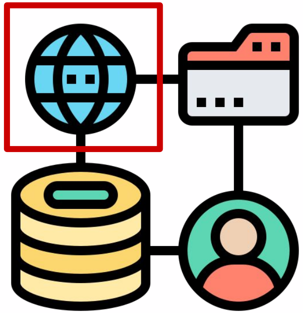
Materials and methods: We performed LC-OCT imaging of PP lesions from 17 patients before and after 8 weeks of treatment. The clinical severity of the single lesions was evaluated using a lesion score (LS), designed considering three parameters: erythema, desquamation, and infiltration. LC-OCT images were segmented by artificial intelligence and evaluated based on three microscopic criteria: the thickness of the stratum corneum, the thickness of the living epidermis, and the undulation of the dermo-epidermal junction.

Results: LC-OCT digital analysis allowed recognition and quantification of the three microscopic criteria, showing a reduction of all these during the follow-up. Furthermore, a high correlation between change in LS and the thickness of the stratum corneum and the thickness of the living epidermis was found.

Conclusion: LC-OCT can non-invasively monitor the response of PP to different treatments. Morphometric changes occurring in the psoriatic lesion during the 8-week treatment period were identified by in vivo LC-OCT and measured by using artificial intelligence. Although future studies are required, based on these preliminary results, LC-OCT may represent a valid potential tool for precise monitoring of therapeutic response.

- What metadata is needed?
- Only abstracts are required
- Document has multiple sections
  - Do you keep the structure?

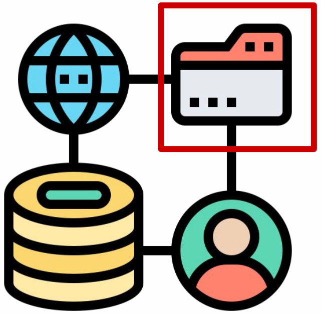
How to Build such System? (Step 2)



- Create a system that considers the entire document set or a small relevant subset to generate the correct answer.
- Mainly two stages
  - **Retrieval**: find documents that are relevant for the question.
  - **Answer generation**: generate an answer based on the selected subset.

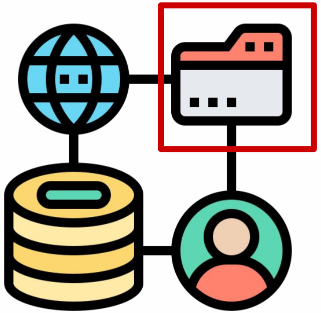
You will learn the necessary tools for this step throughout the course.

How to Build such System? (Step 3)



- Build an interface for the user to connect with the system.
- Command line in simplest form
- Website, mobile app…. (be creative!)

How to Build such System? (Step 4)

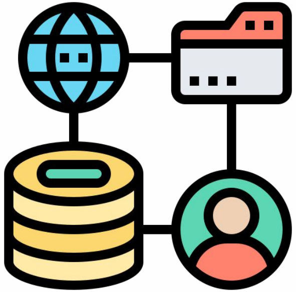


- Come up with an evaluation strategy
  - Make use of manual annotation.
  - Find an automated annotation strategy.
  - Use correct metrics.

- You learn about evaluation metrics in the course

- Take care of edge cases and know the limitations of your system.

Project Timeline



- By 23.10.2023 groups should have been formed
  - Let us know your groups by sending an email to gertz@informatik.uni-heidelberg.de
  - Include for each member: last name, firstname, matriculation number, and program of study
- By 01.11.2023 - mentors are assigned to groups
- 11.01.2024 - First milestone: retrieval and data analysis
  - Contact your mentor and setup a meeting
  - Mentor will / need to talk to all team members
  - Describe what you have done so far
  - GitHub should reflect what you present

Project Timeline /2

- 04.03.2024, 2pm, project deadline
  - Any commit after this time is ignored.

Project grade is based your **code, documentation, report,** and **communication**.

- Comment your code.
- Write installation guide.
- Create a suitable interface.
- Write a report.

Repository Setup

- Everything is contained in a private GitHub repo.
- Create a repo with a meaningful name of the pattern *{project-name}-INLPT-WS2023*.
- Create a meeting folder named **Meetings** and keep meeting notes.
- All necessary information to run the project should be in a **Readme.md** file.
- Download link to the data or description to create the **exact** same dataset should be included in **Readme.md** file.

Report

- Includes the contribution of each team member.
  - We will check the commits as well!

- Add a **documentation.md** file to the project and write the report.
- It should include (template provided later in the course):
  - Motivation
  - Problem definition
  - Model architecture/s
  - Data description and acquisition
  - Evaluation and implementation
  - Conclusions and open issues

Computational Resources

- We recommend that you use open source or cloud-based systems for storage.

- Computations that require GPUs → use Google Colab.

- We do not offer any compute resources.

- No matter what technology you use, make sure that it is easily replicable, we cannot grade what we cannot run!

The Mentors

- Lead Mentor: Ashish Chouhan
- Satya Almasian, John Ziegler, Jayson Salazar, Nicolas Reuter
- Email: *{lastname}*@informatik.uni-heidelberg.de

Communication with Mentors

- Reach out to us when there is a question!
- A mandatory meeting for the first milestone (11.01.2024)
- You can arrange additional meetings if you need guidance.
- Main connection is through the GitHub repository
  - Create a private git repo on GitHub
  - Add your mentor as collaborator
  - Repo must have a meeting logbook

Final Notes

- Start early!
  - The projects are complex and have many parts.
  - There are assignments, classes and procrastination.
- Sometimes the easiest solution is the best solution
  - No need to use the most advanced and complicated method.
  - Study the data and see what fits **your** problem.
- Track your progress
  - Add meeting notes to the github repo.
  - Write down what you work on and read up.
- Start simple and iterate
  - Start with a small prototype, you can make adjustments as you iterate.

Questions?