

SEARCH PROFILES

Iterata Research Platform for identifying and validating patient cohorts based on a document archive system among different hospitals in Switzerland.

Novel Search Approach for Cohort Preselection, Identification, Validation and Structuring

Introduction

Iterata works with a holistic approach where the currently available technical aids should be integrated into daily life. Why doing something just because it has been done for a long time the same way if there is a newer, more efficient way? The connection between human and technical knowledge leads to an improved way of working. Under the principle of *computational thinking mindset*, our Search Profiles are a new possibility for research in every field with big data. The approach is to make big data addressable and with this searchable. The cohort identification and in a second step validation and structuring of the data is much more facilitated. After a short introduction of the application to a domain user, no further need of an IT expert is necessary to perform search requests. The clinical patient reports and laboratory values of the Kantonsspital Baselland, the Kantonsspital Aarau, Bellinzona and St. Gallen are already accessible and addressable. Further approaches will be the collaboration among the different sites.

Preselection

To narrow down the amount of data to further analyze, we propose a preselection with simple regulators. Some generic information about individuals can be defined before the actual search commences. This information would typically be gender, age range, amount of examinations, range within the documents were generated and also the frequency of documents. This preselection allows us to define a smaller cohort with only patients who meet the criteria. The preselection is connected to the document archive or the laboratory values. Only the selected cohort is further analyzed with for example a tally function. This means, that for the preselected cohort, a count is possible which shows how frequently an element (e.g. lab value) is.

Cohort Identification

As a first step, an appropriate and detailed research question is very important. Based on the research question, search terms can be formed. Without knowing what you are looking for, the search request won't be sufficient. The search terms need to be put in the right order, which is in a staircase shape, where the term at the top has the highest priority. Like that, columns are linked with an AND linkage, cells directly below each other are linked with an OR linkage. The first count is based on security level 2-3, meaning the names of the patients are not shown. The result is a count of the hit rate per documents per request (spreadsheet row) and the number of patients. In general, Iterata search possibilities will support strategically two ways. One way is to find the best, adequate and manageable patient selection throughout a hospital or in multi-centers. Second approach is not overlook any patient to certain research question. After the identification phase, the cohort set will be introduced to the Domain or Business Owner in regards to decide about the next action and sample set for validation.

Cohort Validation

The validation is split along data & result governance into level 3 (speed-validation) and level 4 access (detail validation), where in a first step only the hit counts compared to the search terms per patients are validated without further information about the patient or the documents behind the terms. Patients that need a closer look can be validated on level 4 (information about name, birth date and sex), if the researcher has the appropriate access level. There a list of documents behind every search request and a pdf viewer, where the documents can be validated in more detail.

Cohort Structuring

As a completion, the result can be structured in a table form where the number of hit counts for each individual is shown. This is necessary to identify best matches for the search criteria and research question, but also for further statistical analysis in general.

Next Avenues

This approach can not only facilitate the cohort identification in general by making the documents accessible and searchable but gives the opportunity to even find a cohort that is big enough especially in rare diseases. The goal is to build up a network between different hospital sites to connect the research and identify patients of interest. All of this based on a security access system, that only the data is shown that the researcher has access for.

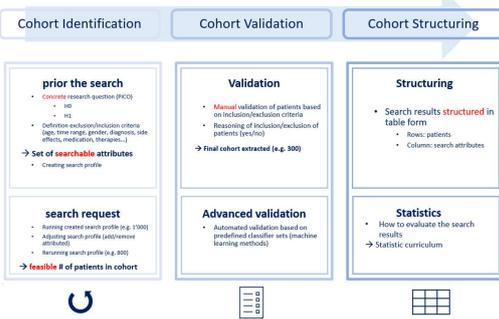
- Collaboration of different hospitals in Switzerland
- Expansion to different hospital sites in Europe to develop an international research collaboration (Horizon 2020)
- Facilitating searches in banking systems and insurances

Please do not hesitate to contact us

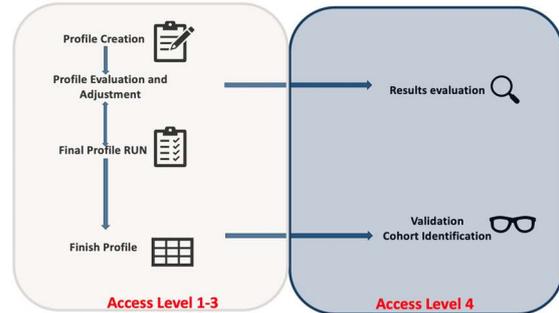
Sincerely yours, Iterata Team

Phone +41 62 842 88 27 | info@iterata.ch

Insights & Impressions



3 schematic Phases from Identification, Validation to Structuring



Search Profile from Level 3 (anonymous) to 4 (depersonalized)

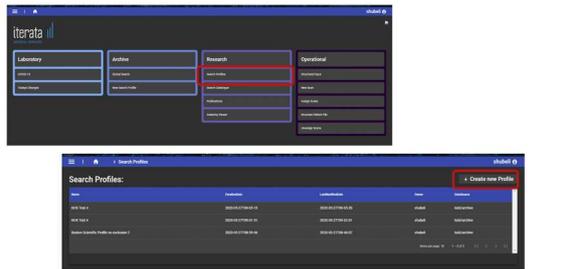
Item #	Item A	Item B	Item C	Item D
Item 1	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 2	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 3	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 4	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 5	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 6	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 7	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 8	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 9	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 10	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 11	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 12	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 13	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 14	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 15	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 16	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 17	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 18	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 19	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus
Item 20	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus	Diabetes mellitus

Semantic for Excel Tree check in:

- AND- An AND Linkage is represented in a row (horizontal merges), exact phrases between ... Each term is in the same row in the 1:n columns
- OR- An OR Linkage is represented in combination of more than one row (vertical merges)
- OR in cell: [Item, Item, ..., Item] = [Item OR Item OR ... OR Item]
- NDT: is represented as: [Item or phrases]
- End-with: [Item]; Start-with: [Item]; Substring: [Item, phrases]

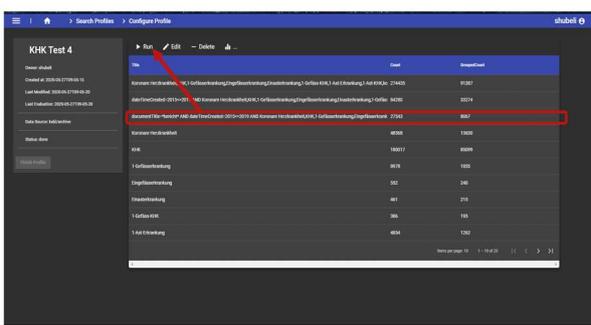
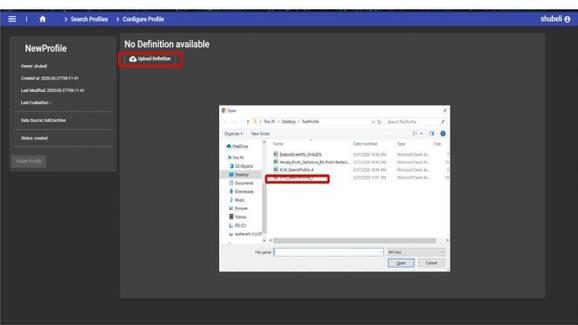
All searches e.g. cell items or phrases + will be executed as an exact term search, but with - at the end a Proximity search will be performed [Item Item]-

Fuzzy: just per item could be weighted with a fuzzy factor, e.g. spelling mistakes "Auchglin", "Vorhofflimm"



Search Profile Creation

Example Front Page Research Platform



Search Profile (Spreadsheet) upload

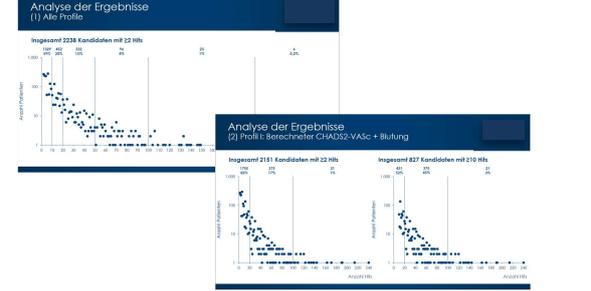
SearchProfile run: Output Hit counts per search row of the spreadsheet

Details zu den Daten

- Zeitraum: zwei Jahre (2018 & 2019)
- Kandidaten mit VHF: 8734
- LAAC Kandidaten: 2417 (27%)
- Kandidaten mit zwei Hits in einer Abfrage werden als relevant angesehen

Kandidaten nach Profil*

- Profil I:** Berechneter CHADS2-VASc + Blutung
 → 1356 Suchprofile
 → 2151 Kandidaten **Option 1 Validierung**
- Profil II:** Berechneter CHADS2-VASc + HAS-BLED Score
 → 160 Suchprofile
 → 68 Kandidaten
- Profil III:** CHADS2-VASc + HAS-BLED Score
 → 56 Suchprofile
 → 40 Kandidaten
- Profil IV:** CHADS2-VASc + Blutung
 → 465 Suchprofile
 → 390 Kandidaten **Option 2 Validierung**



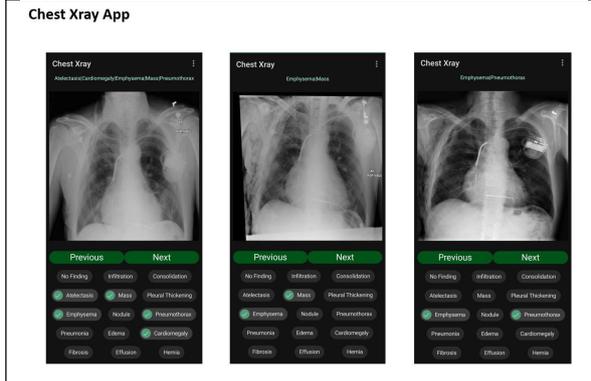
Result Presentation for Decision Making [Examples, Boston Scientific]

Decision and Selection of Cohort



Level 3 Speed Validation (Anonymus)

Level 4 Detail Validation

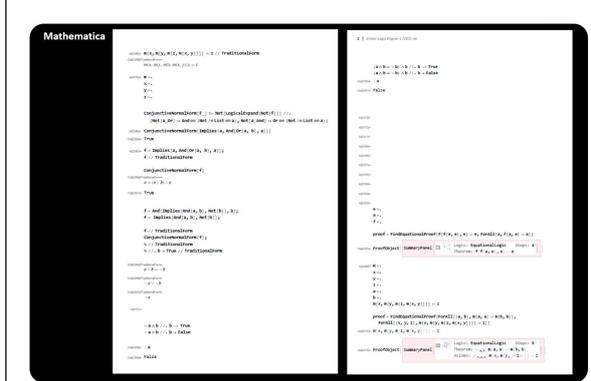


Level 3 Speed Validation on mobile, tablet

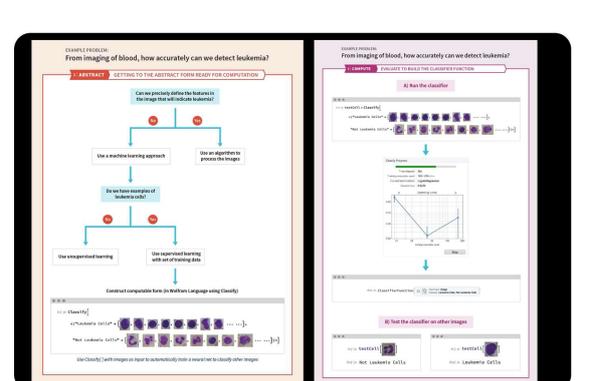
CHEST X-RAY APP

Chief Complaint	Count	Chief Complaint	Count
No Finding	68361	No Finding	68361
Infiltration	19494	Infiltration	9547
Effusion	13317	Atelectasis	4215
Atelectasis	11559	Effusion	3055
Nodule	6331	Nodule	2785
Mass	5782	Pneumothorax	2194
Pneumothorax	5302	Mass	2139
Consolidation	4667	Effusion/Infiltration	1683
Pleural Thickening	3385	Atelectasis/Infiltration	1358
Cardiomegaly	2776	Consolidation	1318
Emphysema	2516	Atelectasis/Effusion	1165
Edema	2383	Pleural Thickening	1136
Fibrosis	1686	Cardiomegaly	8093
Pneumonia	1431	Emphysema	892
Hernia	227	Infiltration/Nodule	829
		Atelectasis/Effusion/Infiltration	737
		Fibrosis	727
		Edema	628
		Cardiomegaly/Effusion	484
		Consolidation/Infiltration	441
		Infiltration/Mass	420
		Effusion/Pneumothorax	403
		Effusion/Mass	402
		Atelectasis/Consolidation	398
		Edema/Infiltration	392
		Infiltration/Pneumothorax	345
		Emphysema/Pneumothorax	337
		Consolidation/Effusion	337
		Pneumonia	322

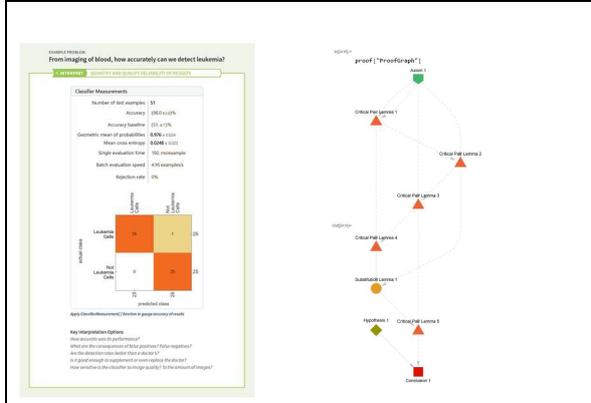
All combination of NHI Chest X-Ray data



Logic Engine – Classifier



Classifier – Expert Trained Set (Neuronal Network)



Classifier Measurement