

Introducing Electronic Health Records to Automate Medical Research

F Bennaoui*, Z Chabihi, N Elidrissi Slitine and FMR Maoulainine

Neonatal Intensive Care Unit, CHU MED VI, Marrakech, Research Laboratory: Children, Health, and Development, FMPM, Cadi Ayyad University, Morocco

***Corresponding author:** F Bennaoui, Neonatal Intensive Care Unit, CHU MED VI, Marrakech, Research Laboratory: Children, Health, and Development, FMPM, Cadi Ayyad University, Morocco

ARTICLE INFO

Received: 📅 March 07, 2024

Published: 📅 March 18, 2024

Citation: F Bennaoui, Z Chabihi, N Elidrissi Slitine and FMR Maoulainine. Introducing Electronic Health Records to Automate Medical Research. Biomed J Sci & Tech Res 55(4)-2024. BJSTR. MS.ID.008724.

ABSTRACT

Every medical research makes use and is centered on data, particularly patient record data; these data are then explored, queried, visualized, and used to draw new conclusions; Data is an essential aspect of any research, yet automating the data life cycle helps shorten the research life-cycle and thus helps establish a continuous research pipeline. To automate the data life-cycle, it is mandatory to define data quality metrics to assert quality guarantees; those assertions can then be identified by assessing data quality dimensions. Current medical record formats, which are mainly paper-based format, are proven to suffer from free-entry system drawbacks. In addition to being prone to error, it is maybe also unreadable. Our objective was to develop a system with four major containers: Hospital management system, practice management system, electronic medical record, and a clinical data warehouse. We used "meta" suite to create a multi-purpose content management system and a universal data management system, and we've encoded content schemes to instruct visual components. The resulting system consists of 8 major containers: hospital management system, practice management system, electronic medical record, multi-purpose content management system, universal data management system, identity, and access management system, internationalization system in addition to the ontology container. We've developed an ontology container to augment our entry system capabilities, which represent concepts in a tree-like structure; each concept is encoded with a prefix string following a Trie data structure, which enables static operations in One time; We've added to the database 270,516 concepts for anatomy, diagnoses, findings, interventions, procedures, medications, organisms, substances in addition to dozen other attributes. The software was thoroughly tested using unit tests, integration tests, and acceptance tests; we've conducted an acceptance end-test: we've submitted 19 real medical records picked form two months period, and we've conducted data exploitation and analysis trial. Our solution can be considered as Big Data solution as it asserts all its attributes; Automation has a beneficial impact on research and healthcare and can be applied through many research processes: Data quality and governance, Headline suggestion, data querying and exploration, reports writing, citations management, and research quality. Major automation frameworks are data quality, artificial intelligence, and natural language processing.

Keywords: Data; Electronic Health; Systeme

Introduction

As the scientific enterprise has grown and diverse, we need empirical evidence on the research process to test and apply interventions that make it more efficient and its results more reliable. Meta-research is the study of research using research methods. Also known as «research on research» aims to reduce waste and increase research quality in all fields. Meta-research concerns itself with research efficiency improvement and the detection of bias, methodological flaws, and other errors and inefficiencies [1]. Medical

research (or biomedical research), also known as experimental medicine, encompasses a wide array of research, extending from fundamental research – involving fundamental scientific principles that may apply to a preclinical understanding – to clinical research, which involves studies of people who may be subjects in clinical trials [2]. Medical research often involves fundamental sciences such as mathematics, physics, chemistry, and philosophy throughout its lifecycle and as a paradigm. Medical research typology is either interventional or non-interventional research; Research is said to be interventional if it interferes with patient management or requires

an additional or unusual monitoring or diagnostic procedure. Interventional research involves biomedical research and healthcare interventions; as of non-interventional research, it often applies statistical inference, in the form of prospective, retrospective, and clinical essay trials [3].

Each type of typologies is conducted upon data collected or centered on patients, which puts data collection and the quality of data at the utmost priority of every medical research; it also determines the quality of the research, and how long it would take to complete, as well as how much it would cost (financial spending) [4]. Health informatics is the interdisciplinary study of the design, development, adoption, and application of IT-based innovations in health care services delivery, management, and planning. Health care informatics includes sub-fields of clinical informatics, such as pathology informatics, clinical research informatics, imaging informatics, public health informatics, community health informatics, home health informatics, nursing informatics, medical informatics, consumer health informatics, clinical bioinformatics, and informatics for education and research in health and medicine, pharmacy informatics. Several systems exist that can be classified as clinical informatics; these systems interleave with each other in terms of concern; these systems help improve patients care by digitizing paper trails, and thus they offer more transparency, legibility, portability, accessibility as well as it can assert certain data quality guarantees depending on the implementation. Health information systems are available to and accessed by healthcare professionals. These include those who deal directly with patients, clinicians, and public health officials. Healthcare professionals collect data and compile it to make health care decisions for individual clients, client groups, and the public.

Objectives

Automating clinical research is a multi-faceted task; most importantly, we aimed to shorten the data lifecycle by applying data governance guidelines and asserting data quality guarantees. A continuous clinical research pipeline would result in shorter iteration cycles, lower financial budgets, higher outcomes, faster feedback loops, and lower churn ratios.

System Containers: The system would eventually consist of these major containers:

- Electronic Health Record System.
- Clinical Data Warehouse.
- Hospital Management System.
- Practice Management System.

Continuous clinical research requires more than basic data governance infrastructure; it would also require multilingual capabilities (mainly English publishing ability), publishing features, security assertions, and conformity to local laws and regulations.

Key Objectives: The key objectives for such systems would be:

- Shorten the lifecycle of clinical research.
- Minimal data capture duration.
- Accurate data capture.
- Increase clinical research outcome.
- Ensure data quality guarantees.
- Offer a friction-less experience to lower the barrier of entry for researchers.
- Ensure the ease of data exploitation. Democratize and secure access to clinical data.
- Anonymize access to clinical data.
- Lower the linguistic overhead and offer a multilingual experience.
- Ensure conformance with personal data protection regulations and laws.
- Being able to conduct highly customizable data queries.
- Multiple charting and data visualization solutions.
- Being exhaustive in clinical data capture.
- Schema flexibility, given that clinical data is always subject to updates.

Materials and Methods

Software Development Lifecycle

Systems development life cycle (SDLC), also referred to as the application development life cycle, is a process for planning, creating, testing, and deploying an information system. There are usually six stages in this cycle: requirement analysis, design, development and testing, implementation, documentation, and evaluation.

Agile Methodology

In software development, agile approaches development requirements and solutions through the collaborative effort of self-organizing and cross-functional teams and their stakeholders. It advocates adaptive planning, evolutionary development, early delivery, and continual improvement, encouraging flexible responses to change.

Results

Visualizing a software architecture and decomposing it into components can be tedious; the C4 model is the most used model to visualize hierarchies in software engineering. The C4 model is an «abstraction-first» approach to diagramming software architecture, based upon abstractions that reflect how software architects and developers think about and build software. The small set of abstractions and diagram types makes the C4 model easy to learn and use.

Level 1: System Diagram

A System Context diagram is a good starting point for diagramming and documenting a software system, allowing you to step back and see the big picture. Draw a diagram showing your system as a box in the center, surrounded by its users and the other systems it interacts with.

- Core Context: Universal content management system (CMS), which theoretically can be tailored to any need and accommodate any schema.
- Major Contexts: include hospital management system, practice management system, electronic health record, and clinical data warehouse contexts.
- Person: the intended audience is the whole medical and paramedical staff.
- Technology: the technologies used are mainly web technologies, following the latest trends, we used Typescript – a

statically type-checked superset of JavaScript – in both backend (through NodeJS), and frontend development, NPM (Node Package Manager) was used to manage external dependencies (Table 1 & Figure 1).

Table 1: Major application contexts.

Context	Description
HMS context	the system must offer HMS capabilities such as HR management, procedure management, and other administrative procedures
PMS context	users of the system must be able to carry their practice routines
EHR context	patient charts, procedures, interventions, prescriptions, and all their admission data must be recorded and displayed in highly accessible interfaces and/or format
CDW context	clinical data must be centralized, and its quality assessed, governed, and easily accessed through secure channels, and can be transformed and queried at the real time



Figure 1: Level 1 c4 diagram applied to our system.

Level 2: Container Diagram

The Container diagram shows the high-level shape of the software architecture and how responsibilities are distributed across it. It also shows the major technological choices and how the containers communicate with one another. It's a simple, high-level technology focused diagram that is useful for software developers and support/operations staff alike.

A. Parsing Container

Parsing, syntax analysis, or syntactic analysis is the process of

analyzing a string of symbols, either in natural language, computer languages, or data structures, conforming to the rules of formal grammar. Parsing consists of many steps before obtaining the abstract syntax tree, these steps are tokenization, lexing, parsing, desugarifying. The parsing container is the core of every operation in our application; it is used to parse datatypes, expressions, schemes, and language templates. Parsing syntax is an intensely repetitive task, and therefore it must be optimized, cached, and predictable. Our parser is a top-down LL parser-combinators parser that syntactically analyzes expressions using custom grammar trees (Figure 2).

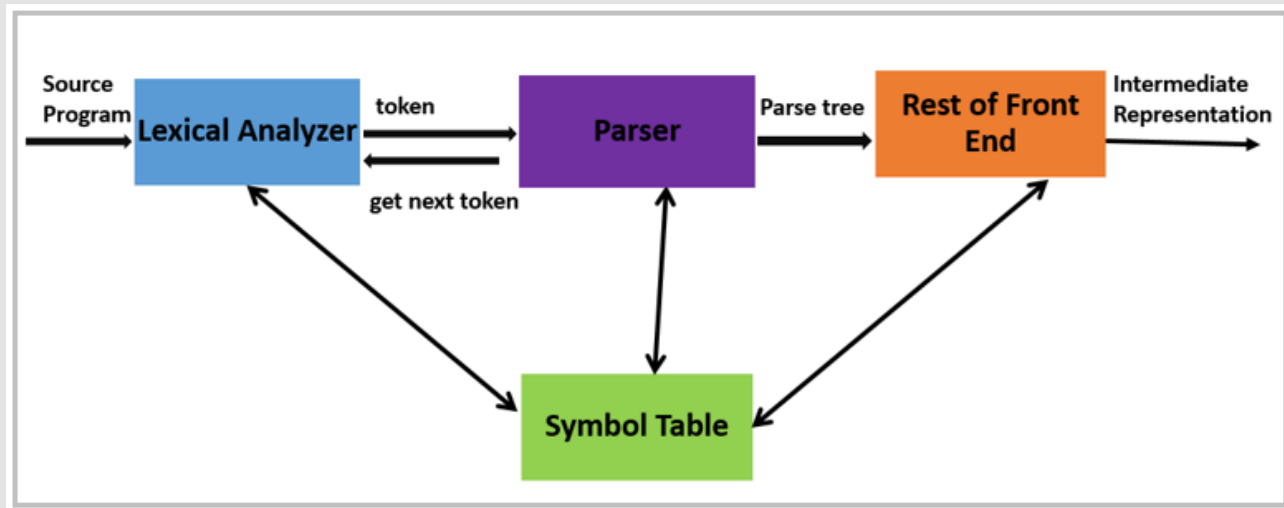


Figure 2: Parsing container and its layers of abstraction.

B. Datatype Container

A data type or simply type is an attribute of data that tells the compiler or interpreter how the programmer intends to use the data. A data type constrains the values that expression, such as a variable or a function, might take. This data type defines the operations that can be done on the data, the meaning of the data, and how that type can be stored. A data type provides a set of values from which an expression (i.e., variable, function, etc.) may take its values. Datatypes are an essential concept in our application; datatypes are the cornerstone of expressions, segments, and conditions. Datatype engine was supplied by metatype, which is a proprietary datatype engine that parses datatypes, validates data against datatypes, generate fake realistic data based on datatypes, generate empty deterministic placeholders, can be used to infer, and apply arithmetics on datatypes among other features (Figure 3). It comes with native support for 22 primitive types, with support for compound types such as arrays and objects, and merge types such as union and intersection. Primitive types are extended with custom attributes and pattern names and flags with over 40 attributes and 60 patterns supported. For example, to come

up with a datatype that matches emails between 10 and 60 characters, one can use "string (min=10, max=60) as email". It supports type coercion, normalization, sanitization, synonym down-sampling, and fault tolerance through custom tolerance flags. Type coercion stands for type conversion capability on inter-coercible types, such as the coercion of a string of digits into an integer. Data normalization or canonicalization is the operation of normalizing similar data into a lookalike pattern such as lower-casing all letters; normalization can be contextual e.g., name cases must not be normalized, whereas usernames must be all lower-cased. Synonym downsampling is the process of normalizing synonyms of the same concept into one concept, e.g., "true," "yes," "on" and "valid" would be downsampled into TRUE Boolean value. Sanitization is the process of stripping out unwanted characters to clean up data, e.g., removing white spaces out passwords, emails, or usernames. Type arithmetic consists of union and intersection merges and subset and superset checks. Also, it supports file manipulation, including file trans-coding, file validation, and asynchronous file transformation and coercion of image, video, audio, and document files. Every feature of metatype has been concisely used in our application; for example, type arithmetic's and

type checking were used in the inference engine and the validation engine respectively of the Expression container. File manipulation capabilities of metatype were mandatory to allow file upload. Metatype also allows access to datatype metadata by accessing raw

parsed data; this property is potent to generate input and form interface in complete synchrony with the datatype counterpart in real time.

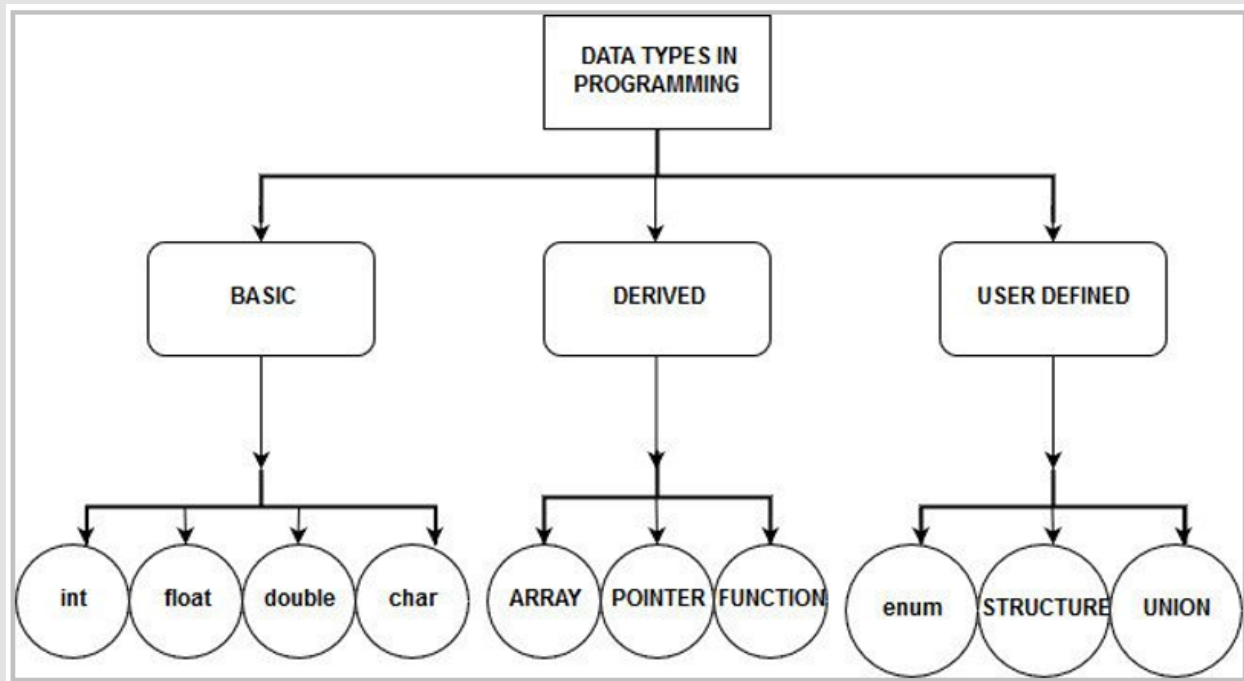


Figure 3: Basic datatypes in programming.

Discussion

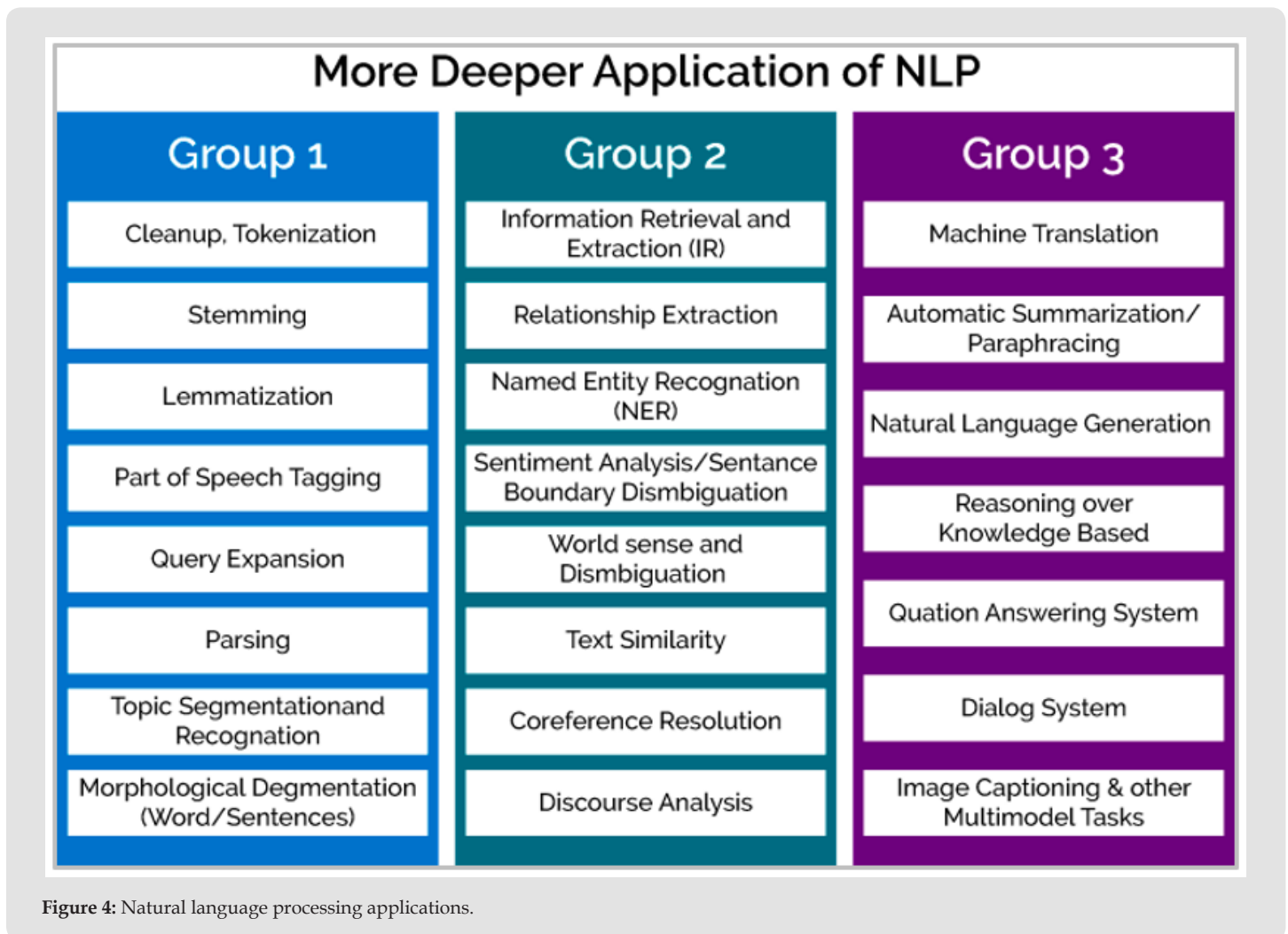
In academic publishing, a paper is an academic work that is usually published in an academic journal. It contains original research results or reviews of existing results. Such a paper is also called an article [5]. The production process, controlled by a production editor or publisher, then takes an article through copy editing, typesetting, inclusion in a specific issue of a journal, and then printing and online publication. Academic copy editing seeks to ensure that an article conforms to the journal's house style, that all of the referencing and labeling are correct, and that the text is consistent and legible; often, this work involves substantive editing and negotiating with the authors [6]. The author will review and correct proofs at one or more stages in the production process. The proof correction cycle has historically been labor-intensive as handwritten comments by authors and editors are manually transcribed by a proofreader onto a clean version of the proof. In the early 21st century, this process was streamlined by the introduction of e-annotations in Microsoft Word, Adobe Acrobat, and other programs, but it remained a time-consuming and error-prone process. The full automation of the proof correction cycles has only become possible with the onset of

online collaborative writing platforms, such as Authorea, Google Docs, and various others, where a remote service oversees the copy-editing interactions of multiple authors and exposes them as explicit, actionable historical events [7].

Several technologies exist that can automate the process report writing, Document formatting standards like LaTeX or markdown that enable creating standardized rich documents using plain text instructions. Markdown is a lightweight markup language with plain-text-formatting syntax, created in 2004 by John Gruber with Aaron Swartz. Markdown is often used for formatting readme files, for writing messages in online discussion forums, and to create rich text using a plain text editor [8]. LaTeX is a software system for document preparation. When writing, the writer uses plain text instead of the formatted text found in «What You See Is What You Get» word processors like Microsoft Word, LibreOffice Writer, and Apple Pages. The writer uses markup tagging conventions to define the general structure of a document (such as article, book, and letter), to stylize text throughout a document (such as bold and italics), and to add citations and cross-references [9]. Markdown documents allow the automated extraction of headlines, tables, figures; tables and figures

can be auto-incremented automatically according to any numbering format; table of contents can be automatically generated with links to the corresponding paragraph. LaTeX can display any mathematical formula, in addition to its being the publishing standard for many journals. Natural language processing algorithms can be used to

write paragraphs, summarize headlines, abstracts, detect plagiarism, speech tone, voice, or compute metrics such as reading time, reading difficulty, word count, or auto-complete sentence, spell check, suggest synonyms, simplify phrases ... etc (Figure 4).



Citations Management

Enumerative bibliographies are based on a unifying principle such as creator, subject, date, topic, or other characteristics. An entry in an enumerative bibliography provides the core elements of a text resource, including a title, the creator(s), publication date, and place of publication [10]. A bibliography may be arranged by the author, topic, or some other scheme. Annotated bibliographies give descriptions about how each source is useful to an author in constructing a paper or argument. These descriptions, usually a few sentences long, summarize the source and describe its relevance. Reference management software may be used to track references and generate

bibliographies as required [10]. Reference management software, citation management software, or bibliographic management software is software for scholars and authors to use for recording and utilizing bibliographic citations (references) and managing project references either as a company or an individual. Once a citation has been recorded, it can be used repeatedly to generate bibliographies, such as lists of references in scholarly books, articles, and essays. The development of reference management packages has been driven by the rapid expansion of scientific literature [10]. These software packages typically consist of a database in which full bibliographic references can be entered, plus a system for generating selective lists of articles in the different formats required by publishers and

scholarly journals. Modern reference management packages can usually be integrated with word processors so that a reference list in the appropriate format is produced automatically as an article is written, reducing the risk that a cited source is not included in the reference list. They will also have a facility for importing the details of publications from bibliographic databases [10].

Research Quality and Peer-Reviewing

Quality research most commonly denotes the scientific process, including all aspects of study design; in particular, it relates to the judgment regarding the match between the methods and questions, selection of subjects, measurement of outcomes, and protection against systematic bias, nonsystematic bias, and inferential error. Principles and standards for quality research designs are commonly found in texts, reports, essays, and guides to research design and methodology, and so on [11]. Besides, quality assessment plays a vital role in the research community. It enlightens crucial decisions on the funding of projects, teams, and whole institutions, on how research is conducted, on recruitment and promotion, on what is published or disseminated, and on what researchers and others choose to read. It makes trust in the work of the research community. Quality is, of course, not a straightforward concept. The Oxford English Dictionary (OED) defines it as the nature or standard of something as measured against other things of a similar kind, and especially the degree of excellence it possesses [11]. Research does investigate ideas and uncovers useful knowledge. But research can be abused through bad assessing of research work. An assessment process implies a review – involving human judgments and/or quantitative scores – which may find work of varying quality, from the poor or mediocre to the excellent or upstanding. So, there are guidelines for standards for research quality [11].

Standards for Assessing the Quality of Research

- Pose a significant, important question that can be investigated empirically, and that contributes to the knowledge base.
- A well-defined research topic and a clear hypothesis.
- Test questions that are linked to relevant theory.
- Apply methods that best address the research questions of interest.
- Base research on transparent chains of inferential reasoning supported and justified by complete coverage of the relevant literature.
- Provide the necessary information to reproduce or replicate the study.
- Ensure the study design, methods, and procedures are sufficiently transparent and ensure an independent, balanced, and objective approach to the research.
- Provide a sufficient description of the sample, the intervention, and any comparison groups.

- Use appropriate and reliable conceptualization and measurement of variables.
- Evaluate alternative explanations for any findings.
- High-quality data were fit for their intended use and reliable, valid, relevant, and accurate.
- Findings of the study written in a way which brings clarity to important issues.
- Tables and graphics which are clear, accurate, and understandable with appropriate labeling of data values, cut points, and thresholds.
- Include both statistical significance results and effect sizes when possible.
- The conclusions and recommendations both logical and consistent with the findings.
- Assess the possible impact of systematic bias.
- Submit research to a peer-review process.
- Adhere to quality standards for reporting (i.e., clear, cogent, complete).
- Is respectful to people with other perspectives.
- Provides adequate references.
- Attempts to honestly present all perspectives.

Natural-language processing algorithms can be used to compute metrics such as voice, tone, formality, conciseness; NLP algorithms can detect plagiarism, classify deception, spot fraud, detected incited sentences, and detect stealing. Citations can be checked for their style, their publisher's impact factor that can be crossed over some predefined threshold ... etc.

Overlooked Features for Assessing the Quality of Research [12]

- Research questions are designed to reach a particular conclusion.
- Alternative perspectives or contrary findings are ignored or suppressed.
- Data and analysis methods are biased.
- Conclusions are based on faulty logic.
- Limitations of analysis are ignored, and the implications of results are exaggerated.
- Critical data and analysis details are unavailable for review by others.
- Researchers are unqualified and unfamiliar with specialized issues.
- Citations are primarily from special interest groups or popular media, rather than from peer-reviewed professional and academic organizations.

Conclusion

Every medical research makes use and is centered on data, particularly patient record data; these data are then explored, queried, visualized, and used to draw new conclusions; Data is an essential aspect of any research, yet automating the data life cycle helps shorten the research life cycle and thus helps establish a continuous research pipeline. To automate the data life cycle, it is mandatory to define data quality metrics to assert quality guarantees; those assertions can then be identified by accessing data quality dimensions. Current medical record formats, which are mainly paper-based format, are proven to suffer from free-entry system drawbacks. In addition to being prone to error, it is maybe also unreadable. Our objective was to develop a system with four major containers: Hospital management system, practice management system, electronic medical record, and a clinical data warehouse. We used "meta" suite to create a multi-purpose content management system and a universal data management system, and we've encoded content schemes to instruct visual components. The resulting system consists of 8 major containers: hospital management system, practice management system, electronic medical record, multi-purpose content management system, universal data management system, identity, and access management system, internationalization system in addition to the ontology container.

We've developed an ontology container to augment our entry system capabilities, which represent concepts in a tree-like structure; each concept is encoded with a prefix string following a Trie data structure, which enables static operations in one time; We've added to the database 270,516 concepts for anatomy, diagnoses, findings, interventions, procedures, medications, organisms, substances in addition to dozen other attributes. The software was thoroughly tested using unit tests, integration tests, and acceptance tests; we've conducted an acceptance end-test: we've submitted 19 real medical records picked form two months period, and we've conducted data exploitation and analysis trial. Our solution can be considered as

Big Data solution as it asserts all its attributes; Automation has a beneficial impact on research and healthcare and can be applied through many research processes: Data quality and governance, Headline suggestion, data querying and exploration, reports writing, citations management, and research quality. Major automation frameworks are data quality, artificial intelligence, and natural language processing.

Conflict of Interest

No.

References

1. JPA Ioannidis, D Fanelli, DD Dunne, SN Goodman (2015) Meta-research: Evaluation and Improvement of Research Methods and Practices. PLOS Biology 13: 10.
2. (2013) AAMC. "Basic Science," Tomorrow's Doctors, Tomorrow's Cures.
3. B Röhrig, JB Du Prel, D Wachtlin, M Blettner (2009) Types of study in medical research: part 3 of a series on evaluation of scientific publications. Deutsches Arzteblatt International 106(5): 262-268.
4. L Cai, Y Zhu (2015) The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal 4: 2.
5. V Larivière, S Haustein, P Mongeon (2015) The Oligopoly of Academic Publishers in the Digital Era. PLOS ONE 10(6): 1-15.
6. AE Jinha (2010) Article 50 million: an estimate of the number of scholarly articles in existence. Learned Publishing 23(3): 258-263.
7. (2016) IETF. The text/markdown Media Type - RFC 7763." Internet Engineering Task Force (IETF).
8. Lesli Lammport (1986) LATEX: a document preparation system. Reading Mass: Addison Wesley Pub Co.
9. B Fredson (1971) Four Faces of Bibliography. Papers of the Bibliographical Society of Canada.
10. (2018) Wikipedia Definition. Reference software. Wikipedia.
11. INNSPUB. Standards of a Quality Research.
12. PM Wortman (1994) Judging research quality. The handbook of research synthesis, pp. 97-109.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2024.55.008724

F Bennaoui. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>