v1: 11 December 2023

Research Article

Kalya Research: Complementary and Alternative Medicine (CAM) Virtual Research Assistant From Biomedical Literature

Peer-approved: 11 December 2023

Qeios, Vol. 5 (2023)
ISSN: 2632-3834

Jessica Pinaire[1], Jean-Marc Durand[1], Philippe Lenoir[1], Frédéric Assié[2], Waleed Ragheb[3], Loric Rivière[1], Guillaume Soulié[1], Anthony Fraise[1]

1. Independent researcher; 2. French National Centre for Scientific Research, Paris, France; 3. Department of Information Technology, Faculty of Computers and Artificial Intelligences, Cairo University, Cairo, Egypt

Complementary and alternative medicines (CAM) have become an emerging subject of interest both for users and health professionals. Rigorous studies identify efficient and safe methods for human health, frequently called by researchers non-pharmacological interventions. The challenge is to determine relevant articles in a large and increasing volume of publications and journals. To meet this challenge, we created Kalya Research (KR), a medical assistant tool based on artificial intelligence that selects and characterizes CAM literature and brings support to medical researchers. Based on rule models and ontologies, KR can suggest relevant and recent CAM publications. It presents key indicators through analytical visualizations. KR was evaluated at several points (effectiveness, relevance, usability) in 2 ways: by means of a bibliographic search comparison with MedLine and by questioning more than 40 biomedical researchers who used KR for their research. When compared with Medline, KR highlighted most of the relevant CAM publications. The evaluation by the researchers showed that the majority of them found the tool to be relevant, a time saver, and feature-rich. Our future objectives are therefore to constantly develop the application to improve our models for detecting CAM publications and named entities (diseases, CAMs, outcomes), and to extend it to new health topics.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

Complementary and alternative medicines and non-pharmacological interventions have increased extensively in biomedical literature toward an integrative approach to medicine and health. Patient demands and practitioner preferences motivated researchers to intensify studies on this topic. An Australian survey evaluated the practice of complementary medicine (CM) by general practitioners (GP) and concluded that there is a real need for evidence-based CM[1]. In parallel, a Chinese review evaluated the quality of reporting Randomized Clinical Trials (RCTs) on traditional Chinese medicine treatments[2]. The review also concluded a real need for improving the standards of RCT reports. A more recent study performed a bibliometric analysis of apitherapy from the scientific literature covering a 36-year period[3]. This study revealed that although there is a growing interest in apitherapy, an imbalance in research on this subject is observed in the literature, which could be explained by the lack of resources. These different examples illustrate the point and show the diversity of therapies used in CM. Knowledge in the biomedical field is advancing every day, reaching millions of publications. For example, in cancer, which represents one of the leading causes of death in the world after cardiovascular disease[4], there are approximately 6.6 million publications (GoogleScholar November 2023). Among them, Kalya assessed that about 600,000 (10%) studies focus on complementary and alternative medicines (CAM). Besides, this volume increased by 58% during the last decade. Thus, the problem is how to identify these publications when there is no repository.

Besides, one of the difficulties that researchers are facing is the exploitation of a vast volume of information resulting from a given request. Converting this mass of information into a structured form is a major challenge that constitutes the starting point for the development and fine-tuning of a suitable query and automatic processing tool. To meet these needs, we created a virtual research assistant based on text mining techniques and specific ontologies to find, sort, and analyze worldwide CAM scientific publications. Currently, we have essentially elaborated this tool for the topic of cancer, but extensions to other health topics are planned and still under development. In this article, we present Kalya Research (KR) cancer prevention and care as a real-time digital system dedicated to evidence-based CAMs. It references all CAM that are evaluated from clinical trials, reviews, systematic reviews, and meta-analyses, which are also published in peer-reviewed scientific journals.

KR identifies all the elements contained in a publication and characterizes these elements by creating mastered metadata and ontologies. One of KR's strengths is its ability to characterize the trio: outcomes, diseases, and CAMs. With KR, researchers can refine their results with original filters and thus better target the desired corpus of articles. Not only can they also visualize correlations and trends, but they can also visualize themes that allow them to nourish their reflection. In addition, it can also visualize the strongholds of its research. Furthermore, KR has been designed to stay as close as possible to the needs of its users. With this in mind, we compared our tool with a Medline search for CAM publications addressing alopecia issues in women with breast cancer.

The aim of this paper is to present KR, our new bibliographic research assistant dedicated to CAM. In the following, we provide a state-of-the-art review of existing bibliographic search tools in section 2. Next, we introduce the KR tool, its functionality, and its use in section 3. Then, we present the application architecture and implementation in section 4. In addition, we explain the experimental comparison between KR and Medline in a concrete example of bibliographic research, as well as the impressions of test users. In section 5, we present the results of this experiment. Section 6 is dedicated to summarizing the current approach, discussing the lessons learned, and suggesting future directions to meet researchers' expectations. Finally, we conclude the paper in section 0.

2. Related Work

With the quickly increasing volume of biomedical literature, any biomedical researcher finds themselves confronted with obstacles well known to scientists, namely the extraction of relevant information, the sorting of documents, and their exploration. To overcome these obstacles, there are already various tools like text exploration by visual analytics that allow filtering query results[5]. The functionalities differ according to the tools. Some of them allow querying directly using keywords and obtaining a list of publications with the annotated abstract on Named Entity Recognition (NER). This is the case for BioIE (based on a rule model)[6], LitSense (specialized in sentence retrieval)[7], and GeneView (targeted biological entities), which also annotates the full text if it is available[8]. Other tools provide additional functionalities, as in Thalia, which displays the frequencies of an entity in the corpus of texts resulting from the query[9], BioTextQuest+ that suggests various methods to cluster the abstracts[10]. Some tools provide other functionalities such as BEST[11], which is a biomedical entity search tool that provides visual analytics on frequent terms and an interaction network for each identified entity. There is also FACTA+[12], which furnishes the biomedical associated concepts with some text analysis pipeline. There are some tools dedicated to a specific task like Quertle[13], which performs a semantic search in multiple biomedical databases (PubMed included) and runs a query via relationships between concepts. BioReader[14] is a binary text clustering tool enabling filtering relevant articles from query results. Then, in a general context,

doi.org/10.32388/IW54X7.2

we notice that text mining techniques are becoming essential tools for finding and sorting relevant articles.

Digital bibliographic tools dedicated to CAMs are also available[15]. Table 1 lists and compares CAM databases. CAM-QUEST®[16], MOTRIAL[17], LIVIVO[18], AMED[19], and CAMbase[20] are bibliographic databases covering a wide spectrum of CAM therapies. Only CAM-QUEST has a pre-defined request system by disease, therapy, and study design categories. CAMbase has not been kept up-to-date since 2005, and AMED groups mainly European journals. Moreover, there are many specialized bibliographic databases, for example, PEDro for physiotherapy[21], HOMEOINDEX for homeopathy[22], ABIM for phytotherapy and Indian medicine[23], NAPRALERT for natural products[24], OTseeker for occupational therapy[25], Arthedata for art therapy[26], CAIRSS for music therapy[27], and CARDS for dietary supplements[28]. Furthermore, the enthusiasm of patients for natural medicines such as Chinese medicine questions "evidence-based medicine," and the number of websites dedicated to it is significant compared to other disciplines. The tendency to evaluate these practices is also growing, and the bibliographic databases are flourishing. Many examples, including - but not limited to - AcuTrials[29], MANTIS[30], CNKI[31], Societas Medicinae Sinensis[32], Wanfang[33] and Qigong and Energy Medicine Database[34]. Thus, by targeting these bibliographic tools, it is possible to query the different databases and extract relevant information in several languages on a specific subject. Then, the next task would be to collect all the query results from the same search on different search engines to sort. That will not be an easy task; to the best of our knowledge, there is not any automatic tool to perform this task. In addition, tasks such as identifying the most influential authors on a subject, cross-checking information (e.g., a disease with a specific therapy and a given outcome), NER, relationship extraction, or topic modelling are less obvious. Therefore, we observe a lack in this area to make an easy search with just a few clicks. To fill this gap, we created KR.

Name	Description	Subject focus	Strengths	Weaknesses
ABIM[23]	Bibliography of India Medicine	Indian Medicine	• Specialized Focus • Comprehensive Collection • Enhanced Research • Search Functionality	• Scope and Coverage • Data Maintenance and Updates (Last on 2015)
AcuTrials[29]	The database is compiled primarily from PubMed, the Cochrane Library and the OCOM library.	Acupuncture	• Specialized Focus on Acupuncture • Search Functionality	• Data Maintenance and Updates (Last on 2018) • Limited Scope or Coverage
AMED[19]	It covers three separate subject areas: professions allied to medicine, including physiotherapy, occupational therapy, rehabilitation, speech and language therapy, and podiatry; complementary medicine; and palliative care.	All CAM	• Comprehensive Coverage • Diverse Content • Evidence-Based Information • User-Friendly Interface	• Limited Full-Text Access: • Scope and Currency • Quality Control
Arthedata[26]	Art therapy-specific bibliographic database, set up in March 2008	Art therapy	• Specialized Focus on Art Data • Comprehensive Collection	• Language Barrier • Limited Scope or Coverage
CAIRSS[27]	CAIRSS (Computer-Assisted Information Retrieval Service System) for Music is a bibliographic database for music research literature.	Music therapy	• Specialized Focus on Music Research Literature • Comprehensive Coverage	• Outdated • Scope and Inclusivity • Data Quality and Relevance
CAMbase[20]	CAM bibliographic database	All CAM	• Evidence-Based Information • Multidisciplinary Content	Data Maintenance and Updates (Last on 2005)
CAM Cancer	Publications are reviewed before selection with strict guidelines	Cancer	• Evidence-Based Information • Multidisciplinary Approach	• Search Functionality • Limited Scope

doi.org/10.32388/IW54X7.2

Name	Description	Subject focus	Strengths	Weaknesses
CAM-QUEST[16]	Information on complementary and alternative medicine	acupuncture, anthroposophic medicine, ayurveda, bioenergetics, homeopathy, manual medicine, mind-body medicine, herbal medicine and TCM	• Search Functionality • Extensive Coverage • German search engine	• Language Barrier • Data Maintenance and Updates (Last on 2019)
CARDS[28]	Computer Access to Research on Dietary Supplements	Dietary supplements	• Search Functionality • Graphical Visualisation	• Limited Scope: • Language and Content Accessibility
CNKI[31]	China National Knowledge Infrastructure Database	TCM	• Recent Publications • Extensive Coverage • Chinese Language Resources	Language Barrier
FACTA+[12]	Text-mining system that helps users explore various types of (possibly hidden) associations in an easy and comprehensible manner		Search Functionality	• Not targeted on CAM • Data Maintenance and Updates (Last on 2019)
HOMEOINDEX[22]	The methods adopted are those used by major international medical databases (MEDLINE, LILACS), allowing data exchange.	Homeopathy	• Multidisciplinary Coverage • Comprehensive Search Functionality • Integration of Multiple Sources	• Limited Full-Text Access • Search Interface Complexity • Coverage Variability • Technical Issues and Updates
ISCMR	Members publications on Traditional, Complementary, and Alternative Medicine	All CAM	• Scientific Rigor • Comprehensive Coverage	• Search Functionality • Limited to members publications

Name	Description	Subject focus	Strengths	Weaknesses
LIVIVO[18]	Interdisciplinary search engine for literature and information in the field of life sciences.	Medicine, health, nutrition, and environmental and agricultural sciences	• Multidisciplinary Coverage • Comprehensive Search Functionality • Integration of Multiple Sources	• Not targeted on CAM • Limited Full-Text Access • Search Interface Complexity • Coverage Variability • Technical Issues and Updates
MANTIS[30]	Citations and abstracts from all alternative medicine disciplines.	Acupuncture, Chiropractic, Exercise therapy, herbal medicine, massage, midwifery, moxibustion, oriental traditional medicine, osteopathic medicine, wellness care	• Comprehensive Coverage • Diverse Information • Research-Focused • User-Friendly Interface • Credible Sources	• Coverage Gaps • Variable Quality of Information • Outdated Information
MOTRIAL[17]	Free and academic meta-search engine from data collection of non-pharmacological intervention trials	All CAM	• Evidence-Based Information • Search Functionality • Specialized Focus	Language Barrier
NAPRALET[24]	Natural Product Alert is a relational database of natural products, including ethnomedical information, pharmacological/biochemical information on extracts of organisms in vitro, in situ, in vivo, in human (case reports, non-clinical trials) and clinical studies.	Natural products	• Extensive Database • Scientific Rigor • Comprehensive Coverage • Search Functionality	• Data Accuracy • Complexity for Lay Users
NCCIH[35]		All CAM	• Evidence-Based Information • Diverse Topics	• Limited Scope • Changing Nature of Research • Potential Bias

doi.org/10.32388/IW54X7.2

Name	Description	Subject focus	Strengths	Weaknesses
NORPHCAM[36]	The Network of Researchers in the Public Health of Complementary and Alternative Medicine is an international collaborative network dedicated to promoting and advancing the public health and health services research of traditional, complementary and alternative medicine and integrative health care.	All CAM	• Comprehensive Coverage • Research Reference • Support for Evidence-Based Practices	• Updates and Maintenance (Last in 2008) • Cross query not available
OTseeker[25]	Referenced trials have been critically appraised for their validity and interpretability.	Occupational therapy	• Specialized Focus • Structured Search System • Evidence-Based Practice Support • Quality Assessment	• Limited Scope • Availability of Full Text • Updates and Maintenance (Last in 2016)
PEDro[21]	Only, randomised controlled trials, systematic reviews and evidence-based clinical practice guidelines are indexed	Physiotherapy	• Focused on Physiotherapy and Rehabilitation • Structured Rating System • Quality-Controlled Content • User-Friendly Interface	• Cross research query not available. • Limited Scope • Restricted Access to Full Text • Updates and Inclusion of Recent Studies
Qigong and Energy Medicine Database[34]	The Qigong and Energy Medicine Database™ contains abstracts collected by the Qigong Institute since 1984	Qigong and Energy Medicine	• Comprehensive Coverage • Research Reference • Support for Evidence-Based Practices	• Depth of Information • Outdated Content • Quality Assurance
SciELO[37]	Scientific Electronic Library Online is a bibliographic database, digital library, and cooperative electronic publishing model of open access journals	Health information on latin america literature	• Open Access Content • Regional Representation • Multidisciplinary Coverage • Quality Control	• Not targeted on CAM • Language Barrier • Coverage Variability • Limited Global Reach • Indexing and Citations • Technical Infrastructure and Updates

doi.org/10.32388/IW54X7.2

Name	Description	Subject focus	Strengths	Weaknesses
Societas Medicinae Sinensis[32]		TCM	• Extensive Coverage • German Language Resources • Local Perspectives and Research • Diverse Content Types	• Language Barrier • Limited internationality
Wanfang[33]	An essential resource for medical research in China, which covers medical journals, dissertations, conference proceedings, patents, standards, companies and products.	TCM	• Extensive Coverage • Chinese Language Resources • Local Perspectives and Research • Diverse Content Types	• Language Barrier • Quality and Curation • Limited International Coverage

Table 1. Comparing CAM databases strengths and weaknesses

3. Tool Description

Currently, KR references 350,088 publications, 2,220 Non-Pharmacological Interventions (NPI), and 2,050 outcomes related to cancer. These results continue to grow as the tool integrates new publications daily. In this section, we present the usage of the tool through its search engine version and its visualization capabilities.

After login, users access the KR interface. The user interface is divided into three main parts (Figure 6): a search box, the left-side panel to select criteria, and the right-side panel for displaying the results.

In the search box, the user can enter keywords, and by clicking on the “advanced” button below, users can add and/or exclude disease(s), NPI(s), outcome(s), or other terms. In addition, by clicking on the “Save” button, users can save and name their queries and choose to send the results to their mailboxes. In the following, we will describe the tool through a case study research example: what are the NPIs that prevent alopecia following chemotherapy? To answer this question, we enter the term chemotherapy in the search box and click on the advanced button to enter the term alopecia in the outcome research criterion (Figure 1 and Figure 2). We save our query by clicking on the green “save” button (Figure 2) so that we will be notified about new studies on this subject. The number of articles concerned with chemotherapy-induced alopecia (CIA) contained in the KR database is displayed below on the results panel. At the top left of the right-side panel, there are two tabs: “View by analytics” and “View by list.” The first offers an analytical visualization of the results while keeping the filters panel available. The second offers a raw view of the list of articles as presented above. Thus, the first display shows the graphical analysis of the results, and we can see the list of NPIs studied and used to remedy this problem. On the right of the banner, a logout button allows the users to leave the search area.

doi.org/10.32388/IW54X7.2

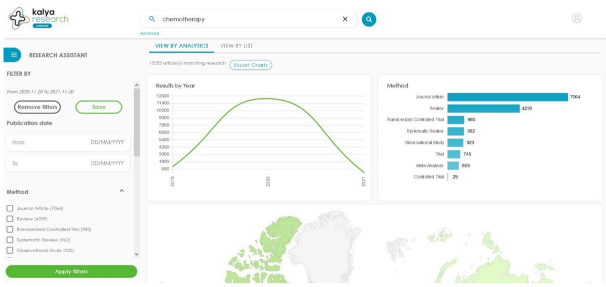

Figure 1. Screenshot of a research example in the Kalya Research tool – enter query.

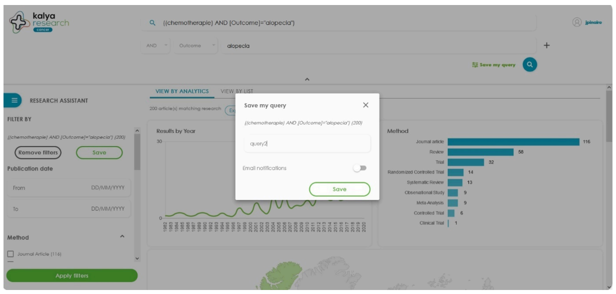

Figure 2. Screenshot of a research example in the Kalya Research tool – save query.

Graph-based visualization

Users can navigate through the selection of articles via different graphics. KR offers an original visualization of the different key indicators such as the number of publications by year or by methods (Figure 1) and the countries interested in these research issues (Figure 3). Figure 4 represents the most cited NPIs; in our previously mentioned example, the most studied is ergonomic tools. There are also the most targeted outcomes, the cancer localization, and the most frequent journals. The graphics are interactive. Indeed, users can also apply filters or refine their results by clicking on the bars of the histograms. For example, a user can choose to focus only on one NPI, the scalp cooling. To do so, users should click on the scalp cooling bar of the NPIs histogram at the top left (Figure 4), which will automatically reduce the selection (Figure 5).

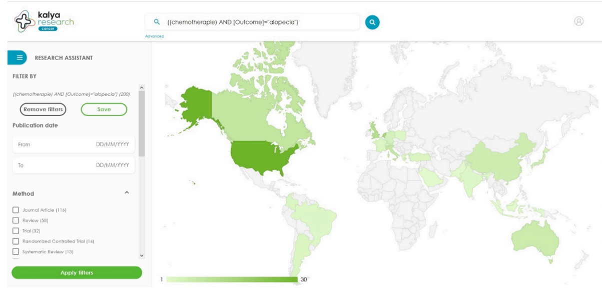

Figure 3. World map of articles related to chemotherapy-induced alopecia.

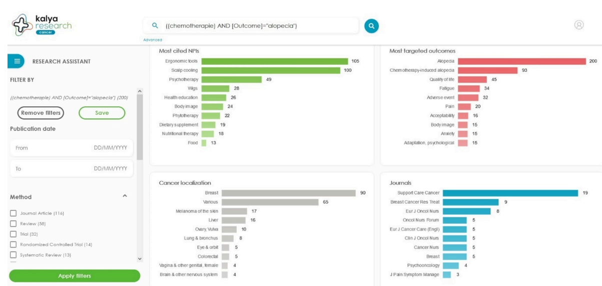

Figure 4. Visualizations of key indicators summarizing results of the query related to chemotherapy-induced alopecia.

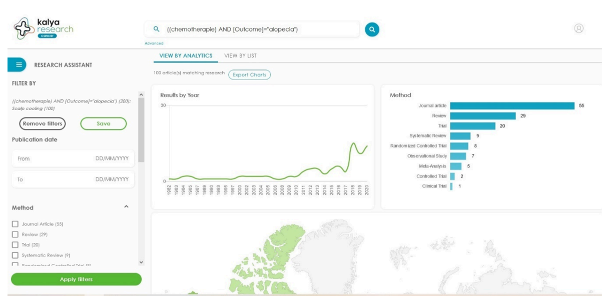

Figure 5. Refine the query from chemotherapy-induced alopecia to scalp cooling NPI by clicking directly on the graphs.

Text information

In the “View by List” panel, the list of articles is displayed under the search box in the right-side panel, in chronological order - from the most recent to the oldest. For each article, KR displays its title, its journal name, its date of publication, and the name of the first author. By clicking on the black arrow, the first lines of the abstract appear, as well as the health goal, the disease, the outcomes, and NPI terms related to this article. In our example (Figure 7), for the first article, the health goal is Care and the disease is Various cancer (which means all cancers), whereas the outcomes are alopecia and fatigue, while the NPI terms related to the article are the physical activity. The hand display, allowing users to click and access the article reference page, is activated by hovering over the article information.

The left-side panel allows the user to add filters to the initial search (Figure 6 and Figure 7). The publication period can be defined by filling in the fields under Publication date. Then, the study design of the article sought can be defined by checking the corresponding boxes in the "Methods" field. For example, it is easy to quickly know the number of trials (Figure 3). The choice is made by checking among the following methods: journal article, review, trial, observational study, randomized controlled trial (RCT), controlled trial, a meta-analysis (MA), and systematic review (SR). In the same way, the symptoms sought can be selected in the "Outcomes" field located just below. The list includes all outcomes related to the list of articles resulting from the initial research (made using the search box). It is the same for the "Disease" field. The health goal criterion includes the following choices: Care, Cure, and Prevent. Indeed, Care is related to supportive or treatment care, Cure is related to healing or recovery, and Prevent is related to health troubles that require prevention. The different types of the population studied are Adult, Child, and Elderly. The minimum number of participants is chosen by filling the ad hoc input field. Next, a selection can be made on the country using the checkbox. Finally, the last criterion concerns the Research design of the study with the following choices: Mechanistic, Interventional, Observational, Implementation, and Prototypal. In detail, a mechanistic study uses various experiments to find causal models (e.g., in vitro experiments), an interventional study compares benefits and risks of NPIs on several outcomes, an observational study describes the natural use and response of NPIs, an implementation study is a monitoring study, and a prototypal study is a design study to evaluate the feasibility of a study.

All filters apply when the user clicks on the "Apply filter" button. These filters can be either deleted or saved for a future search by clicking on the relevant buttons ("Remove filters" or "Save") at the top of the panel. In addition, at the top left of the filter panel, a blue button displays a contextual menu offering the user access to the saved queries, as well as to the personal lists of the saved articles.

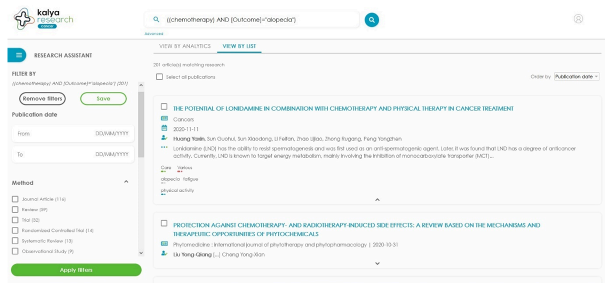

Figure 6. Screenshot of a research example in the Kalya Research tool – View by List - view 1.

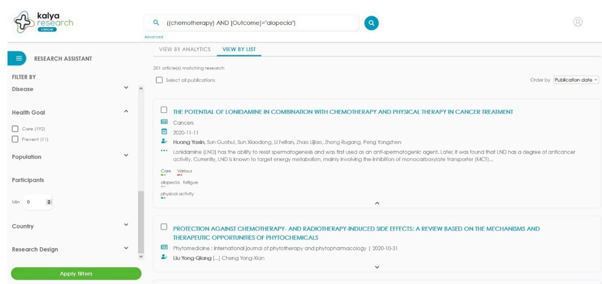

Figure 7. Screenshot of a research example in the Kalya Research tool - View by List – view 2.

Back to our CIA research example, we had selected the NPIs of scalp cooling. To have the best levels of proof, we check RCT, SR, and MA in the Method filter (Figure 8), and if we focus on the most recent publication, i.e., the most recent, we can see by reading the conclusion of the abstract that scalp cooling would be the most effective method to prevent CIA.

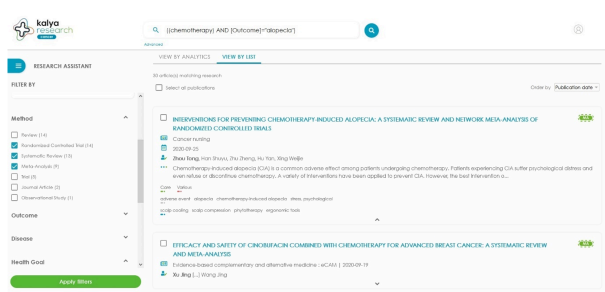

Figure 8. Kalya research tool: refine a query with the Method filter to have the best level of proof.

doi.org/10.32388/IW54X7.2

When there are new results for my saved query, they appear in the contextual menu under My Queries, as shown in the left part of Figure 9. Then, by clicking on new results, we see the list of saved queries, and by clicking again on the request that you want to view, the new articles are displayed on a gray background.

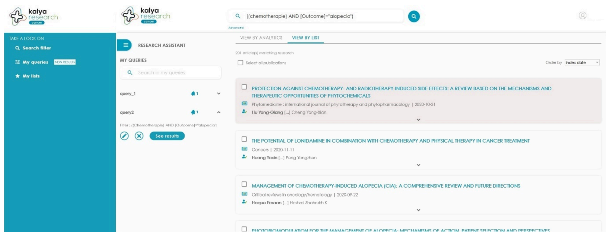

Figure 9. Kalya research tool displays new results for saved queries. By clicking in the contextual menu in the left banner, New Results is displayed in front of My Queries (left part of the Figure, blue background). By clicking on the query concerned, the new articles appear on a gray background (right part of the Figure).

4. Methods

In this section, we present the tool's design, architecture, and implementation. Finally, we describe how our tool compares with a Medline search in a concrete example of bibliographic research.

Material

KR focuses on CAM techniques, also called NPI. "NPIs are non-invasive methods of care (programs, products, or services) whose efficacy in improving the health and quality of life (QOL) of human beings has been proven. Their effects on health and QOL markers are observable (with measured risks and benefits beyond mere user opinions) and can be linked to identified biological and/or psychosocial processes. They can also have a positive impact on health behaviors and socio-economic indicators."[38][39] According to the French Platform CEPS (*Collaborative d'Evaluation des programmes de Prévention et de Soins de support*) platform, NPIs can be divided into 5 categories of intervention: nutritional; digital; physical; psychological; and elemental interventions.

To identify CAM or NPI publications, we have created a lexicon of NPI terms, classified according to the 5 categories specified above. The classification is based on various sources such as the CEPS platform (https://plateformeceps.www.univ-montp3.fr/fr/english-0), the lexical resources from the National Center for Complementary and Integrative Health (NCCIH) [40], and many others. This list cannot be exhaustive because of the daily monitoring of the different resources.

PubMed is considered the most commonly used search engine by medical researchers as it identifies much of the biomedical literature. Therefore, we are detecting CAM publications from this resource. At the beginning of 2020, PubMed referenced more than 30 million citations for biomedical literature from MEDLINE, life science journals, and online books. We supplement this resource with other journals specialized in CAM but not referenced in PubMed, such as the Journal of Client-Centered Nursing Care.

Kalya Research Implementation

KR architecture is based on 5 components: frontend, backend, data science, medical and scientific expertise, and user evaluations (Figure 10). The operating principle is as follows. The backend manages data recovery via APIs and integration into data warehouses. The data science component is composed of many modules that make it possible to extract the information of interest and transform this data into valuable information. The frontend provides the web interface that will be accessible to researchers. The medical expertise ensures the integrity and enrichment of the data and makes it possible to decide in the event of litigation. Finally, user assessment allows KR to provide a service that meets the real needs of the end-user.

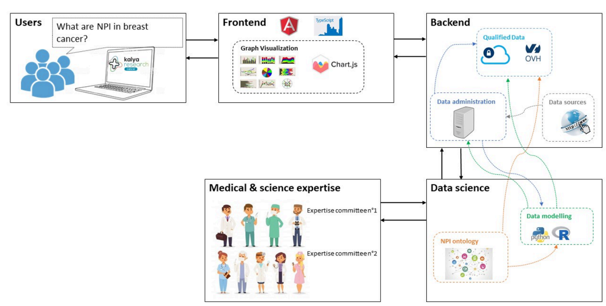

Figure 10. Kalya Research architecture.

The backend component extracts data from biomedical literature with an import script by using APIs. Then, data with a web micro-services architecture which provides an API in .NetCore, to be integrated into a

doi.org/10.32388/IW54X7.2

NoSQL database. This database remains accessible by the data science and frontend components.

The data science component identifies CAM articles with a complex rule-based model briefly explained in the following section.

The frontend component is developed with the IONIC and Angular TypeScript-based open-source front-end web application frameworks for high-performance applications combined with chartJS and Ng2GoogleChartsModule for interactive visualization developments.

The security of the entire architecture is ensured by strong authentication via an Open-Id SSO (Single Sign-On).

Detecting NPI in Natural language

KR uses a rule-based model to identify CAM articles among the thousands of biomedical literature publications. To optimize retrieval effectiveness, we define three rules:

i. The first rule focuses on relevant publications containing NPI terms. Specifically, we define a score based on the term frequency and inverse document frequency (TF-IDF) ranking function to find the most appropriate documents[41]. The score is calculated for a predetermined list of terms: the lexicon of NPI terms, mentioned in the Material section, on a preprocessed text. This text is the result of the concatenation of the title, abstract, and keywords. We use pure text with natural language processing frameworks without lemmatization[42].

1. The second rule focuses on the rejected articles to check for false negatives from the first rule. We define a score based on the occurrences of words combining both the interest terms in addition to the exclusion terms. According to the definition of an NPI, we have created a lexicon of exclusion terms linked, most often, to the administration mode. For example, vitamin C is an NPI, but if the administration mode is by injection, then this method will be rejected, and the article concerned will not be listed in our search engine. The score is applied for the NPI lexicon combined with the exclusion terms lexicon to the same text as for the previous rule, but this time with a lemmatization step.
2. The third rule manages the false positives from the previous two rules. We define a score based on journal relevance and citation occurrences of CAM articles. This score is defined in such a way that it will favor articles published in dedicated journals (with a high CAM score, such as the Integrative Cancer Therapies journal (https://journals.sagepub.com/home/ict)). Besides, articles with many CAM publications in their references. This information is not always available. Indeed, it is possible that a journal has not yet been listed in our model (but listed in our database) or that we do not have access to the reference list of the publication analyzed. In this case, the decision of the last rule takes precedence.

Evaluation of the tool on a practical case

We have evaluated the tool in a practical case, that of alopecia in breast cancer. The aim is to compare the results, their relevance, and the time taken to sort the results.

We used the criteria of the PICO (Patient Intervention Comparison and Outcomes) method. The bibliographic search was carried out using 2 different methods: Medline (see Table 2) and the KR tool (see Table 2). Given the search theme, the criteria are as follows:

- Patient: patients with breast cancer, regardless of age and stage of cancer.
- Intervention: any NPI.
- Comparison: standard care or any NPI.
- Outcome: alopecia.

	Patient	Intervention	Comparison	Outcome
Kalya Research	Breast cancer			alopecia
Medline	("breast cancer"[Title/Abstract] OR "Breast Carcinoma"[Title/Abstract] OR "Mammary Carcinoma"[Title/Abstract] OR "Breast Tumor"[Title/Abstract] OR "Mammary Tumor"[Title/Abstract] OR "Malignant Breast Neoplasm"[Title/Abstract] OR "Invasive Ductal Carcinoma"[Title/Abstract] OR "Invasive Lobular Carcinoma"[Title/Abstract] OR "Ductal Carcinoma in Situ (DCIS)"[Title/Abstract] OR "Lobular Carcinoma in Situ (LCIS)"[Title/Abstract] OR "Triple-Negative Breast Cancer"[Title/Abstract] OR "HER2-Positive Breast Cancer"[Title/Abstract] OR "Estrogen Receptor-Positive Breast Cancer"[Title/Abstract] OR "Progesterone Receptor-Positive Breast Cancer"[Title/Abstract] OR "Metastatic Breast Cancer"[Title/Abstract] OR "Early-Stage Breast Cancer"[Title/Abstract])	(Integrative medicine[MeSH Terms] OR Complementary Therapies[MeSH Terms] OR Alternative Medicine[MeSH Terms] OR Traditional Medicine Mind-Body Therapies[MeSH Terms] OR Dietary Supplements[MeSH Terms] OR Therapeutics[MeSH Terms] OR Physical Therapy Modalities[MeSH Terms] OR Psychotherapy[MeSH Terms] OR Rehabilitation[MeSH Terms]) NOT Tamoxifen NOT "Aromatase inhibitors " NOT Anastrozole NOT Letrozole NOT Exemestane NOT Fulvestrant NOT Palbociclib NOT Ribociclib NOT Adriamycin NOT Doxorubicin NOT Cyclophosphamide NOT Paclitaxel NOT Docetaxel NOT Epirubicin NOT Gemcitabine NOT "5-Fluorouracil 5-FU" NOT Methotrexate NOT Vinorelbine NOT Trastuzumab NOT Herceptin NOT Pertuzumab NOT Perjeta NOT Lapatinib NOT Tykerb NOT T-DM1 NOT "Ado-Trastuzumab Emtansine" NOT Everolimus NOT Afinitor NOT Palbociclib NOT Ibrance NOT Ribociclib NOT Kisqali NOT Abemaciclib NOT Verzenio NOT Atezolizumab NOT Tecentriq NOT Pembrolizumab NOT Keytruda NOT Nivolumab NOT Opdivo NOT Bisphosphonates NOT Denosumab NOT "Zoledronic acid" NOT Zometa NOT Pamidronate NOT Aredia NOT Denosumab NOT Xgeva NOT "PARP Inhibitors" NOT Olaparib NOT Lynparza NOT Talazoparib NOT Talzenna NOT "Cytotoxic Antibiotics" NOT Eribulin NOT Halaven NOT Bevacizumab NOT Avastin		(alopecia[Title/Abstract] or "hair loss"[Title/Abstract])

Table 2. Research algorithms for Kalya Research and Medline for alopecia in breast cancer.

The literature search identified several publications whose quality had yet to be determined. It was therefore necessary to carry out a second selection of the publications thus obtained, so as to optimize the quality of the results and identify the studies answering the research question.

This selection was made based on the title and the abstract. The criteria used to include the articles were the same as those used to define the search strategy.

Rejected studies met one of the following criteria:

- Population: studies not assessing the population concerned, lack of population data
- Intervention: not involving an NMI

doi.org/10.32388/IW54X7.2

- Type of study: pharmacological studies, in vitro or in vivo animal studies, study protocols.

Tool usability assessment

We assessed the usability of the tool with a group of 40 participants. They were mostly medical researchers from France and the USA. The evaluation session was divided into 3 steps: a) presentation of KR; b) training in the basic functionalities of KR; c) collection of user impressions and suggestions to improve KR through open discussions.

5. Results

In this section, we present the results for the comparison with Medline research and the users' feedback on the tool's usability.

Challenging KR with a Medline Research

The results of our literature search for non-drug interventions for alopecia in breast cancer patients are summarized in the flow chart in Figure 11.

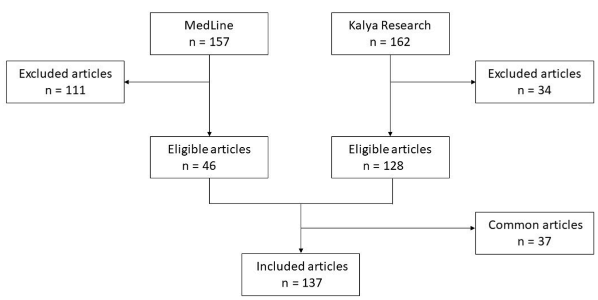

Figure 11. Flow Chart.

We used KR, as described in section 3, filling the search bar with the word 'breast' and the outcome field with 'alopecia' to specify that we wanted publications concerning alopecia in breast cancer. We obtained 162 publications.

In Medline, we used a complex query (see Table 2) with synonyms or related terms to breast cancer, with alopecia and synonyms combined with MeSH terms related to CAM, and excluded a list of drugs approved for the treatment of breast cancer. We obtained 157 publications.

We examined the various publication lists to check that we had publications related to non-drug therapies. The flow chart (see Figure 11) shows that we found more CAM publications with KR than with Medline.

Next, we compared the results and found 37 CAM publications common to both queries. We also noted that 91 publications were in the KR results but not in the Medline results. Conversely, 9 CAM publications were in the Medline results but not in the KR results.

Focus on usability assessment

The discussions with users were essentially focused on three important points: i) the usefulness of the tool, ii) the ease of use and its ergonomics, and iii) the missing elements, the points to be improved.

Of all those surveyed, 80% think KR is useful and interesting; it saves time by having access to the latest biomedical research results on CAM. On the other hand, 20% of the respondents think that everything is already present in PubMed and that this tool is not necessary; everyone is able to make a query with the advanced search form. However, 10% of them were interested in visualizing results as graphs.

The users appreciated the ergonomics, especially on cross-cutting subjects such as QOL and CAM. The features offered were also appreciated. Users found that it is interesting to have access to the various existing CAM used and studied for a given outcome. In addition, KR presents several levels of classification of outcomes and offers the possibility of crossing them. This feature was also appreciated. Besides, the interactive graphics allowing filtering, as well as the readability of the statistics and the convenience of navigation, were valued. Biomedical researchers made few suggestions to improve KR; for instance, some would like more filters to refine their query, for example, the journal name, the impact factor, and the number of citations.

6. Discussion

KR is a new medical research assistant focused on CAM, referencing scientific publications. The originality of KR is to make the data quickly intelligible through visualizations. Regarding the implementation, our choice of tools and methods is simple, classic, and at the same time, it has proven to be very effective [43][44]. This strategy constitutes a solid base to evolve in order to improve our system both in terms of the flexibility of the system to manage the vast increasing amounts of data, as well as the capacity to detect CAM documents. KR provides an even more user-friendly interface for end-users.

The example query presented in this paper shows that KR is more efficient in terms of time, query design, and results obtained. Indeed, while the query can be

doi.org/10.32388/IW54X7.2

formulated quickly and obtained in a few clicks, KR delivers results in a matter of seconds. We have shown that the number of CAM publications is much higher via KR than via Medline. The loss is very small (9 publications found in Medline but not in KR). On the other hand, the development of the query via Medline required a great deal of preliminary thought, necessitating tests (not detailed in this paper). Despite this, the number of CAM articles was lower, and we obtained many off-topic publications. This suggests that the query should certainly be reworked beforehand. However, it appears difficult to compete with a selection made by a learning model process.

KR also competed with Medline in another context. One study presented a systematic literature review of NPIs offered to hormone-suppressed prostate cancer patients according to PRISMA rules, comparing the Medline and KR databases [45]. The authors showed that all publications (n=37) were found by the KR search algorithm, and that the Medline database was able to find 70% of the articles (n=26), but did not uncover any CAM articles different from KR. They concluded that the KR search method was a relevant method for accessing CAM publications.

In parallel, the user evaluation highlighted the usefulness and practicability of KR, contrasted by the lack of time for literature monitoring. This leads to the study of how to further facilitate research to save working time and effort. To remedy this, we plan to integrate text annotation models to facilitate the reading of abstracts by highlighting diseases, outcomes, NPI, etc. We are also studying the possibility of producing a graphical summary of all the abstracts as a complementary result for each query.

In light of these application examples, further improvements are possible. To manage the high volume of data, we planned to strengthen our architecture with Elastic Search for indexing data[46], Redis for the in-memory data structured storage[47], and RabbitMQ to streamline communication between micro-services [48]. Likewise, the NPI detection model is based on methods that have been tested in many natural language processing models[49]. Since we have built a substantial knowledge database, we planned to assess the performance of models via text similarity analysis to detect CAM publications[50]. The challenge is to determine the most appropriate method, in terms of performance in our context, among the different existing approaches. We also plan to refine the annotation of the texts to enrich our metadata system. In the literature, there are many biomedical NER systems linked to the surge of biomedical data[51][52][53]. However, the task of identifying named entities in biomedicine is still a complex task[54]. Another future challenge will be to assess the annotation system that best meets our needs or even combines several ones.

The development of computational linguistics and identification methods in data mining shows promising opportunities to guide the extraction and optimization of the sequestered CAM knowledge. These may be contained within the continuously growing biomedical literature. However, automatic approaches cannot completely replace the in-depth analysis required by a conservative expert. In addition, the diversity of both studies and evaluations challenges the implementation of a robust automatic system. It should also be noted that the construction of the KR database is based on the validation of our experts to reinforce the automatic detection system. Nonetheless, automated systems taken in combination with expert knowledge present much opportunity to leverage new CAM knowledge.

Such a tool as KR is useful for any biomedical researcher in their process of scientific discovery in order to identify a work plan before the start of a scientific study, discuss the results at the end of a study, define the scientific validity of an NPI in the prevention or treatment of several pathologies and health issues (e.g., Nordic walking to combat cancer, meditation to limit stress...), and/or conduct a systematic review more comprehensively. Besides, KR is useful for clinicians looking for new and proven results to support clinical decisions, for instance, to identify which NPIs are the most effective in preventing certain pathologies (Cancer, Diabetes...). The usefulness of KR is even more important as it allows linking diseases with NPIs and their outcomes. Moreover, we have created and structured our metadata in such a way for KR to make it possible to cross queries like "the set of studies on breast cancer and nutritional therapies with an economic evaluation," by using the words "breast" and "nutritional therapies" and checking "eco" in the Outcome field. The user can select the publications based on other outcomes evaluated by the study. We identify several outcomes such as clinical, psychological, social, behavioral, and economic. CAM-QUEST and LIVIVO allow crossing information. However, in CAM-QUEST, the user can filter by study design and no more, whereas in LIVIVO, there are more criteria, but the proposed metadata does not contain the outcomes. So, to the best of our knowledge, this type of search does not exist on any other search engine and accords the uniqueness and novelty of KR.

Limitations

We have identified several limitations to our work, which we describe below.

KR is mainly based on a single data source, PubMed, which is the most common in biomedical research. However, the CAM literature finds its source in lesser-known journals such as Acta Chimica and Pharmaceutica Indica or the Journal of Alternative and Complementary Medicine that are not referenced in PubMed. To remedy this, we could enrich our model with other bibliographic sources (Hindawi, ScienceDirect...).

CAM publications are detected using a single rule model, the operating principle of which is described in the article. However, this strategy has its limitations, as it may miss publications of interest that will not be detected by our model or select publications that will not be of interest. One solution would be to implement several models with different operating modes, such as deep learning models based on BERT classifiers [55][56], and add an arbitration model to select CAM publications. In this way, if several models are matched, we could increase our correct detection rate and decrease our error rate.

While specific approaches to extracting information from text have proved highly effective in the general domain, they have shown their limitations in specific sectors such as healthcare, for several reasons [38]. Firstly, pre-processing chains are difficult to set up to deal with complex medical vocabularies, acronyms, abbreviations, and so on. There is also a variety of expression inherent to each author, which makes it difficult to recognize domain terms in texts and consequently the concepts linked to them, as well as to identify the relationships linking them. Furthermore, the machine learning approaches currently in use generally require a lot of data annotated by human experts, which is difficult to obtain in the medical environment, as annotations require a lot of time - which experts don't have - and sometimes, a significant level of expertise. Labeled examples may contain errors and inconsistencies due to variations in the attention paid by annotators.

What's more, these automatic approaches provide results with little explanation, making them difficult to interpret. In addition, in our case, simple recognition is not enough, as it can lead to confusion. Indeed, if we take the example of vitamin D, it's an NPI when it comes to dietary supplementation, but it's also a blood indicator. We need to find models that take context into account to face this difficulty. Several authors have taken an interest in the subject, one example being the BioALBERT model, which targets the modeling of inter-phrase coherence to better learn context-dependent representations[57].

Finally, there are different terms for the same thing, for example, vitamin B5 and pantothenic acid. The NPI lexicon we've created manages some synonyms, but not all. To remedy this, NPIs' ontologies are under construction[38]. This aims to facilitate the detection of articles, the management of synonyms, and make it possible to propose categories for classifying articles. This can be combined with other sources, for example, iDISK (for dietary supplements)[58], ethnopharmacological knowledge[59], Customary Medicinal Knowledgebase (aboriginal plants)[60], TCMGeneDIT (a database for associated traditional Chinese medicine, gene and disease information using text mining)[61], TM-MC (constituent compounds of medicinal materials in Northeast Asia traditional medicine)[62], and many others. In our perspective, this ontology will be accessible via KR in which the user can navigate.

KR is constantly being updated and enriched with the extension to other medical themes, such as healthy aging and physical activity, which are scheduled for the next version. In addition, technologies are constantly developing to offer an even more efficient and relevant application to meet the needs of our end-users.

7. Conclusion

In this article, we present KR as a novel virtual research assistant tool for biomedical literature. KR is mainly dedicated to CAM. The originality of KR is to offer a double interface allowing end-users to display the list of the results of their queries (as in a conventional search engine), perform an analytic visualization of these results, and notify users with their saved queries' results. Biomedical researchers can easily navigate through these different displays and extract both the publications of interest while scrutinizing key indicators from which they could establish initial findings or orient their research. In the context of enthusiasm for CAM prompted by an awakening of interest in mainstream medicine, an anthology of small tools has emerged[15]. Thus, documentary research on CAM has hitherto been possible using these different search engines but was a real challenge because most of these tools are very specialized for a specific set of therapies, some of which are no longer kept up-to-date [63][64]. KR is therefore prompt and a time saver.

doi.org/10.32388/IW54X7.2

We compared literature searches via KR and Medline to find NPI solutions for breast cancer patients with alopecia. This simple example highlighted KR's advantage over Medline in terms of ease of use, speed, and relevance of results. These initial results are encouraging and open up the prospect of modifications to optimize our detection models.

In addition, KR was evaluated and tested by more than 40 experts. Although our tool has a number of limitations, these evaluations provoke the enthusiasm of the experts on ergonomics and filters in the sense that KR saves months of work. However, the experts believe KR needs to enrich its metadata and to dive into the data by proposing, for a given query, a visual exploration of a text corpus and/or highlighting the key results of this corpus. All these proposed modifications and more come in line with our perspectives for future improvements.

In the future, we plan to improve KR on technical points to streamline data management with the implementation of Elastic Search[46] and RabbitMQ[48] or the implementation of text similarity to detect CAM publications more easily[50]. Besides, we propose to increase KR coverage by adding other biomedical literature sources like Google Scholar and Scopus and by expanding to other health topics. Furthermore, we are working on a new evaluation indicator for scientific publications to measure their scientific relevance and consequently the benefit of a CAM for a given context. In addition, we are working on new functionalities such as NER to highlight diseases, NPIs, and outcomes at a glance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. ^*Pirotta M, Kotsirilos V, Brown J, et al. Complementary Medicine in General Practice: A National Survey of GP Attitudes and Knowledge. Aust Fam Physician 2010; 39: 946.*
2. ^*Li J, Liu Z, Chen R, et al. The quality of reports of randomized clinical trials on traditional Chinese medicine treatments: a systematic review of articles indexed in the China National Knowledge Infrastructure database from 2005 to 2012. BMC Complement Altern Med 2014; 14: 362.*
3. ^*Danell J-AB. Reception of integrative and complementary medicine (ICM) in scientific journals: a citation and co-word analysis. Scientometrics 2014; 98: 807–821.*
4. ^*World Health Organization. Time to deliver: report of the WHO independent high-level commission on noncommunicable diseases. Licence: CC BY-NC-SA 3.0 IGO, Geneva: WHO.*
5. ^*Pinaire J, Azé J, Bringay S, et al. Patient healthcare trajectory. An essential monitoring tool: a systematic review. Health Inf Sci Syst 2017; 5: 1–18.*
6. ^*Divoli A, Attwood TK. BioIE: extracting informative sentences from the biomedical literature. Bioinformatics 2005; 21: 2138–2139.*
7. ^*Allot A, Chen Q, Kim S, et al. LitSense: making sense of biomedical literature at sentence level. Nucleic Acids Res 2019; 47: W594–W599.*
8. ^*Thomas P, Starlinger J, Vowinkel A, et al. GeneView: a comprehensive semantic search engine for PubMed. Nucleic Acids Res 2012; 40: W585–W591.*
9. ^*Soto AJ, Przybyła P, Ananiadou S. Thalia: semantic search engine for biomedical abstracts. Bioinformatics 2019; 35: 1799–1801.*
10. ^*Papanikolaou N, Pavlopoulos GA, Pafilis E, et al. BioTextQuest +: a knowledge integration platform for literature mining and concept discovery. Bioinformatics 2014; 30: 3249–3256.*
11. ^*Lee S, Kim D, Lee K, et al. BEST: Next-Generation Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature. PLoS ONE; 11. Epub ahead of print 19 October 2016. DOI: 10.1371/journal.pone.0164680.*
12. a, b*Tsuruoka Y, Miwa M, Hamamoto K, et al. Discovering and visualizing indirect associations between biomedical concepts. Bioinforma Oxf Engl 2011; 27: i111-119.*
13. ^*Giglia E. Quertle and KNALIJ: searching PubMed has never been so easy and effective. Eur J Phys Rehabil Med 2011; 47: 687–690.*
14. ^*Simon C, Davidsen K, Hansen C, et al. BioReader: a text mining tool for performing classification of biomedical literature. BMC Bioinformatics 2019; 19: 57.*
15. a, b*Boehm K, Raak C, Vollmar HC, et al. An overview of 45 published database resources for complementary and alternative medicine. Health Inf Libr J 2010; 27: 93–105.*
16. a, b*Complementary and alternative medicine: state of clinical research | CAM-Quest, https://www.cam-quest.org/en.*
17. a, b*Ninot G. Motrial - An Academic and Collaborative Search Engine Dedicated to Behavioural and/or Public*

doi.org/10.32388/IW54X7.2

Health Intervention Publications. Eur J Public Health 2 019; 29: ckz185.814.
18. a, b*LIVIVO - The Search Portal for Life Sciences, http s://www.livivo.de/?referer=GREENPILOT/beta2/app (a ccessed 22 June 2020).*
19. a, b*Roberts DJ. AMED: A bibliographic database for co mplementary medicine and allied health. Complemen t Ther Med 1995; 3: 255–258.*
20. a, b*Ostermann T, Zillmann H, Raak CK, et al. CAMbase – A XML-based bibliographical database on Comple mentary and Alternative Medicine (CAM). Biomed Dig it Libr 2007; 4: 2.*
21. a, b*Sherrington C, Herbert RD, Maher CG, et al. PEDro. A database of randomized trials and systematic revie ws in physiotherapy. Man Ther 2000; 5: 223–226.*
22. a, b*Mesquita A, Martins CC, Cepeda LMR. Homeoinde x: New computerized bibliographical database of hom oeopathic literature. Br Homeopath J 1994; 83: 209–21 5.*
23. a, b*School of Electronics and Computer Science. ABIM - An Annotated Bibliography of Indian Medicine. indi anmedicine.nl, http://indianmedicine.eldoc.ub.rug.nl/.*
24. a, b*Loub WD, Farnsworth NR, Soejarto DD, et al. NAPR ALERT: computer handling of natural product researc h data. J Chem Inf Comput Sci 1985; 25: 99–103.*
25. a, b*McKenna K, Bennett S, Dierselhuis Z, et al. Australi an occupational therapists' use of an online evidence-based practice database (OTseeker). Health Inf Libr J 2 005; 22: 205–214.*
26. a, b*Elbing U, Schulze C, Zillmann H, et al. Arthedata—A n online database of scientific references on art therap y. Eur J Integr Med 2009; 1: 39–42.*
27. a, b*Eagle CT, Hodges DA. CAIRSS for music in arts med icine. Int J Arts Med 1992; 1: 2–21.*
28. a, b*Haggans CJ, Regan KS, Brown LM, et al. Computer Access to Research on Dietary Supplements: A Databa se of Federally Funded Dietary Supplement Research. J Nutr 2005; 135: 1796–1799.*
29. a, b*Marx BL, Milley R, Cantor DG, et al. AcuTrials®: an online database of randomized controlled trials and s ystematic reviews of acupuncture. BMC Complement Altern Med 2013; 13: 181.*
30. a, b*Tomasulo P. MANTISTM — Manual, Alternative, an d Natural Therapy Index System Database. Med Ref S erv Q 2001; 20: 45–55.*
31. a, b*Xia J, Wright J, Adams CE. Five large Chinese biome dical bibliographic databases: accessibility and covera ge. Health Inf Libr J 2008; 25: 55–61.*
32. a, b*Fischer T. Der Relaunch der SMS-Literaturdatenba nk zur chinesischen Medizin. Chinesische Med Chin M ed 2016; 31: 41–49.*
33. a, b*Kronenberg F, Molholt P, Zeng ML, et al. A compreh ensive information resource on traditional, compleme ntary, and alternative medicine: toward an internatio nal collaboration. J Altern Complement Med N Y N 20 01; 7: 723–729.*
34. a, b*Sancier KM. Search for medical applications of qig ong with the Qigong Database. J Altern Complement Med N Y N 2001; 7: 93–95.*
35. ^*Ng JY, Dhawan T, Fajardo R-G, et al. The Brief History of Complementary, Alternative, and Integrative Medic ine Terminology and the Development and Creation o f an Operational Definition. Integr Med Res 2023; 12: 1 00978.*
36. ^*David Sibbritt JA. Developing and promoting public h ealth methods for integrative medicine: Examples fro m the field in Australia. J Integr Med 2011; 9: 233–236.*
37. ^*Neves M, Yepes AJ, Névéol A. The Scielo Corpus: a Par allel Corpus of Scientific Publications for Biomedicine. In: Calzolari N, Choukri K, Declerck T, et al. (eds) Proce edings of the Tenth International Conference on Lang uage Resources and Evaluation (LREC'16). Portorož, Sl ovenia: European Language Resources Association (E LRA), pp. 2942–2948.*
38. a, b, c*Nguyen TL, Laurent A, Rapior S, et al. Defining a Collaborative Ontology for Non-Pharmacological Inte rventions, https://hal-lirmm.ccsd.cnrs.fr/lirmm-013831 68 (2016).*
39. ^*Ninot G. Non-Pharmacological Interventions: An Ess ential Answer to Current Demographic, Health, and E nvironmental Transitions. Springer International Publ ishing. Epub ahead of print 2021. DOI: 10.1007/978-3-0 30-60971-9.*
40. ^*Weber WJ, Hopp DC. National Center for Complement ary and Integrative Health Perspectives on Clinical Re search Involving Natural Products. Drug Metab Dispos 2020; 48: 963–965.*
41. ^*Aizawa A. An information-theoretic perspective of tf –idf measures. Inf Process Manag 2003; 39: 45–65.*
42. ^*Indurkhya N, Demerau F. Handbook of Natural Lang uage Processing | Taylor & Francis Group. Chapman a nd Hall/CRC, https://www.taylorfrancis.com/books/97 80429149207 (2010, accessed 29 May 2020).*
43. ^*Stevenson I..Net core architecture. Comput Control E ng 2003; 14: 24–27.*
44. ^*Altinigne CY. Development of a New Web Portal for t he Database on Demand Service. CERN-STUDENTS-N ote-2017-031, Geneva: CERN. IT Department, 2017.*
45. ^*Cecchi M, Ninot G, Rebillard X, et al. Quelles intervent ions non médicamenteuses proposer aux patients trait és par hormono-suppression pour un cancer de la pros*

tate ? Revue systématique de la littérature. Prog En Urol 2023; 33: 287–306.

46. a, b*Kononenko O, Baysal O, Holmes R, et al. Mining modern repositories with elasticsearch. In: Proceedings of the 11th Working Conference on Mining Software Repositories. Hyderabad, India: Association for Computing Machinery, pp. 328–331.*
47. ^*Carlson JL. Redis in Action. USA: Manning Publications Co., 2013.*
48. a, b*Chao-Wei Y, Shu-bo Z. Realization on Asynchronous Full-deplex Message Bus Based on RabbitMQ. Comput Eng Softw 2016; 33.*
49. ^*Bao Y, Deng Z, Wang Y, et al. Using Machine Learning and Natural Language Processing to Review and Classify the Medical Literature on Cancer Susceptibility Genes. JCO Clin Cancer Inform 2019; 1–9.*
50. a, b*H.Gomaa W, A. Fahmy A. A Survey of Text Similarity Approaches. Int J Comput Appl 2013; 68: 13–18.*
51. ^*Miranda-Escalada A, Farré, Eulàlia, Krallinger M. Cantemist corpus: gold standard of oncology clinical cases annotated with CIE-O 3 terminology. 2020; 303–324.*
52. ^*Oliwa T, Maron SB, Chase LM, et al. Obtaining Knowledge in Pathology Reports Through a Natural Language Processing Approach With Classification, Named-Entity Recognition, and Relation-Extraction Heuristics. JCO Clin Cancer Inform 2019; 1–8.*
53. ^*Alshaikhdeeb B, Ahmad K. Biomedical Named Entity Recognition: A Review. Int J Adv Sci Eng Inf Technol 2016; 6: 889.*
54. ^*Campos D, Matos S, Oliveira J, et al. Biomedical named entity recognition: a survey of machine-learning tools. In: Theory and Applications for Advanced Text Mining. InTech Rijeka, Croatia, 2012, pp. 175–195.*
55. ^*Danilov G, Ishankulov T, Kotik K, et al. The Classification of Short Scientific Texts Using Pretrained BERT Model. Public Health Inform 2021; 83–87.*
56. ^*Han Q, Tian S, Zhang J. A PubMedBERT-based Classifier with Data Augmentation Strategy for Detecting Medication Mentions in Tweets. ArXiv211202998 Cs, http://arxiv.org/abs/2112.02998 (2021).*
57. ^*Naseem U, Khushi M, Reddy V, et al. BioALBERT: A Simple and Effective Pre-trained Language Model for Biomedical Named Entity Recognition. In: 2021 International Joint Conference on Neural Networks (IJCNN). 2021, pp. 1–7.*
58. ^*Rizvi RF, Vasilakes J, Adam TJ, et al. iDISK: the integrated DIetary Supplements Knowledge base. J Am Med Inform Assoc 2020; 27: 539–548.*
59. ^*Ningthoujam SS, Talukdar AD, Potsangbam KS, et al. Challenges in developing medicinal plant databases for sharing ethnopharmacological knowledge. J Ethnopharmacol 2012; 141: 9–32.*
60. ^*Gaikwad J, Khanna V, Vemulpad S, et al. CMKb: a web-based prototype for integrating Australian Aboriginal customary medicinal plant knowledge. BMC Bioinformatics 2008; 9: S25.*
61. ^*Fang Y-C, Huang H-C, Chen H-H, et al. TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining. BMC Complement Altern Med 2008; 8: 58.*
62. ^*Kim S-K, Nam S, Jang H, et al. TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. BMC Complement Altern Med 2015; 15: 218.*
63. ^*Shekelle PG, Morton SC, Suttorp MJ, et al. Challenges in Systematic Reviews of Complementary and Alternative Medicine Topics. Ann Intern Med 2005; 142: 1042–1047.*
64. ^*Sarkar IN. Chapter 17 - Challenges in Identification of Potential Phytotherapies from Contemporary Biomedical Literature. In: Mukherjee PK (ed) Evidence-Based Validation of Herbal Medicine. Boston: Elsevier, pp. 363–371.*

Declarations

Funding: This research was entirely funded by Kalya company.
Potential competing interests: No potential competing interests to declare.

 doi.org/10.32388/IW54X7.2