Amman Arab University

7th International Scientific Conference of the Faculty of Business in Partnership with the Deanship of Scientific Research and Graduate Studies

Title: "Digitalization of Business and Scientific Research: Future Visions"

Date : July 27-28, 2024

Thème : Scientific Research and Business Development in the Digital Age

A knowledge bridge: a comprehensive digital database of Arab conferences with Personalized LLM-powered Recommendations

Dr. Friha. Othmane	Pr. Kerouf Mounir	Pr.Friha linda	Dr. Touamria Rym
Environmental Research Center (CRE), Annaba, Algeria	Laboratory of Diversification and Digitization of the Algerian Economy University of 8 May 1945, Algeria	Laboratory of Diversification and Digitization of the Algerian Economy University of 8 May 1945, Algeria	Abbes Laghrour University of Khenchela, Algeria
Friha.othmane@gmail.com	Kharouf.mounir @univ-guelma.dz	Friha.linda@univ guelma.dz	thouamria.rim@univ-khenchela.dz

Abstract:

The number of different specialized conferences has increased, and as such, the appropriate conference-related information with respect to interests for researchers, students, and professionals is hard to find. However, digital databases provide an effective solution by way of a unified platform on which to collect comprehensive information on various conferences. Such a digital database of all Arab conferences would be an important resource for researchers, students, and all those concerned about being updated on the latest developments in many scientific fields. The Arab conference databases provide detailed data and information relating to conferences, such as the title of the conference, date, location, organization, website link, speakers list, program, and working papers.

In addition, and in the continuously growing landscape of scientific research, conferences have become increasingly demanding to navigate. Correspondingly, searching for conferences relevant to specific research agendas can be time-consuming and inefficient. This paper proposes how Large Language Models (LLMs) can revolutionize conference discovery and recommendation for researchers. LLMs excel at Natural Language Processing (NLP) tasks, and we propose leveraging their capabilities to create a personalized conference recommendation system. This system would extract key themes and topics from a researcher publication history and research interests.

Armed with this information, the LLM would search for and recommend conferences dealing with those specific areas. Other helpful services provided in the database may help in one's

search for everything related to subsequent Arab conferences. It also provides information from previous conferences, including papers presented and video recordings of meetings. Further, it is supporting different languages, especially Arabic and English, to spread the scope of access and benefit. Additionally, it creates forums and groups for easing communication between scholars and researchers attending the same conferences, among many other highly regarded services.

1. Introduction:

The increase in more specialized conferences in different scientific fields has been a challenge to the researchers, students, and professionals who may like to get relevant conferences matching their interests. Thus, digital databases took the history of conferences closer to the scientist by unifying details about many important events. In the Arab conferences, it would act as an important resource to ensure updates for the scientific community on various developments. This paper outlines the development of a prototype for a personalized conference recommendation model, utilizing Large Language Models (LLMs) in the pipeline. LLMs have held the state of the art in many Natural Language Processing tasks [1] and proved very useful in extracting semantics from unstructured text. LLM-base recommendation systems has emerged as a hot avenue of research [2]–[4] By making use of LLMs, our system could analyze the publication history of a researcher together with their research interests in order to recommend a conference that best fit their academic pursuits. In Fig. 1 we provide an illustration of our proposed system prototype. As we can see, the user just needs to prompt our system for a recommendation on specific conferences that match its profile.

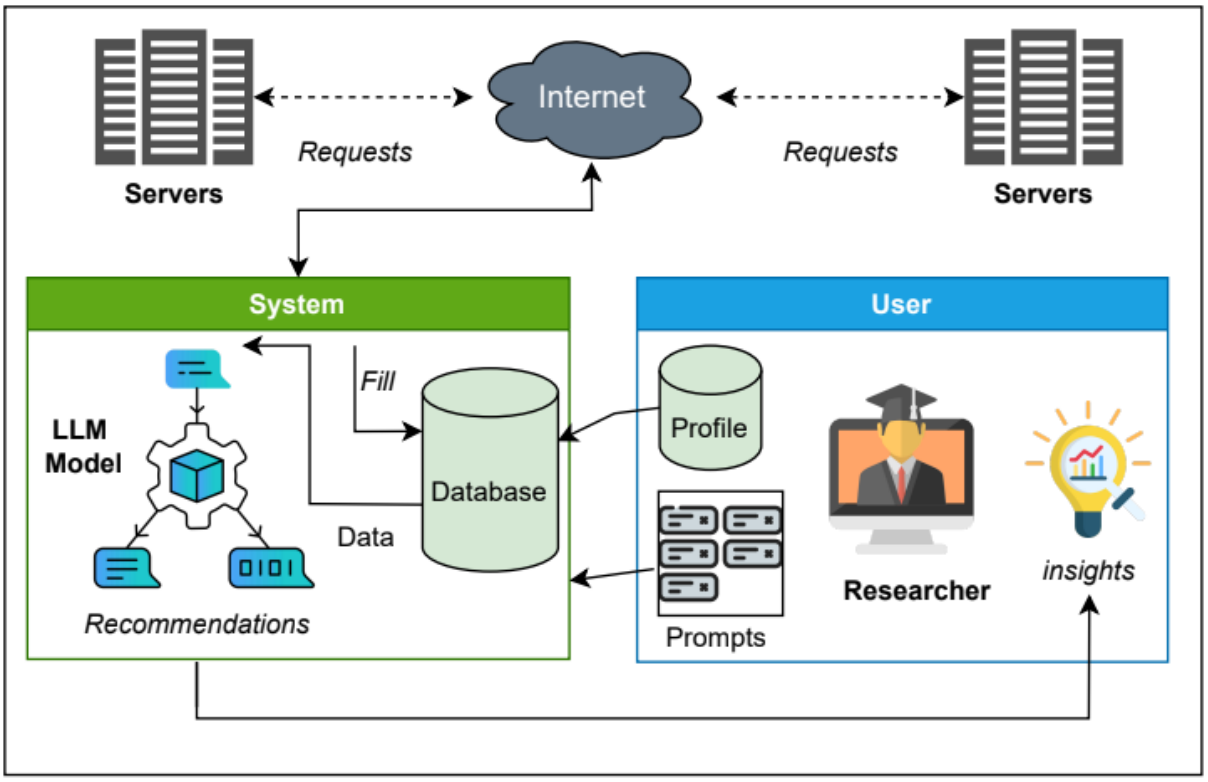

Figure 1 : Our Proposed System Prototype

1.1. Related Works

Various research efforts have examined different methods for conference recommendation systems. For instance, Zhang et al. [5] developed Confer, a tool aimed at helping conference participants find interesting papers and talks, connect with people who have similar interests, and manage their schedules with a personalized conference timetable. Similarly, Li et al. [6] introduced CPRec, a recommendation method that uses both the content and authorship information of papers. This method goes beyond content analysis by incorporating the relationships between users and paper authors into the recommendation process. They extract several features for a user-paper pair from citation networks, co-author networks, and the content itself. Moreover, they derive a user's presences for conference papers based on their bookmarked papers at each conference. Another significant contribution is by Jelodar et al. [7], who applied the Latent Dirichlet Allocation (LDA) technique for conference recommendations. They used probabilistic topic modeling based on Gibbs sampling algorithms to perform semantic mining on eight computer science conference publications from the DBLP dataset. To our knowledge, our research is the first to utilize LLM capabilities in this manner for conference recommendation systems.

2. SYSTEM MODEL

In this part we describe the system model for our contribution. The main blocks for our system include, data collection, profile creation, feature extraction, conference profiling, similarity calculation, and recommendation generation.

2.1. Data Collection

First and foremost, in the development of our personalized recommendation system, we will be gathering comprehensive data related to Arab conferences. The data extends from the title of the conference to the dates of the conference, the location, the organizing body, the link to the website, the speakers list, the program schedule, and working papers. The conference title gives the name of the conference and the dates specify when the conference is held. The location details indicate the venue, whereas the organizing body describes the agency. The website link is provided to have access to more information, while speakers list provides details about keynote and invited speakers. The program contains an outline of the timings of sessions, workshops, and other events, and the Working Papers, with abstracts or the full text of the papers presented. All this information is stored in a digital database supporting efficient querying and retrieval, forming the basis for our recommendation system. This database would need to be implemented so that conferences can be matched with the interests of researchers.

2.2. Profile Creation

We obtain the profile of every researcher, specifying his publication history and research interests. Let ***Pi*** denote the publication history of researcher ***i***, and let ***Ri*** denote the research interests of researcher ***i***. First, it contains a list of papers published by the researcher either solely or co- authored, along with metadata: titles, abstracts, keywords, and publication venues. The research interests, ***Ri***, may either be explicitly provided by the researchers or inferred from their publication history.

2.3. Feature Extraction

Using an LLM, we extract key themes and topics from ***Pi*** and ***Ri***. This process involves generating embeddings [8] for the texts in the publication history and research interests. Let ***Ti*** denote the set of themes and topics extracted for researcher ***i***. Mathematically, we represent the text of the publication history and research interests as vectors in a high-dimensional space. Let ***vp*** be the embedding vector for a publication ***p***, and ***vr*** be the embedding vector for a research interest ***r***. The theme and topic vector ***Ti*** is then given by:

$$T_i = \frac{1}{|P_i| + |R_i|} \left(\sum_{p \in P_i} \mathbf{v}_p + \sum_{r \in R_i} \mathbf{v}_r \right)$$

2.4. Conference Profiling

Similarly, we extract for each conference j in the database the themes and topics ***C***j using the same LLM. All data of the conferences, titles and abstracts of the sessions, and as well as keywords, are transformed into an embedding vector. Let ***vc*** be the vector of the embedding for a conference attribute. The vector of theme and topics ***Cj*** then is:

$$C_j = \frac{1}{|A_j|} \sum_{a \in A_j} \mathbf{v}_c$$

2.5. Similarity Calculation

To match researchers with relevant conferences, we calculate the similarity between the themes and topics of each researcher ***i*** and each conference ***j***. This is achieved using cosine similarity [9]. The cosine similarity ***sim***(***T***i, ***C***j) is defined as:

$$\text{sim}(T_i, C_j) = \frac{T_i \cdot C_j}{\|T_i\| \|C_j\|}$$

where ***Ti*** · ***Cj*** is the dot product of the vectors, and ‖***Ti***‖ and ‖***Cj*** ‖ are the magnitudes of the vectors.

2.6. Recommendation Generation

Based on the similarity scores, we generate a list of recommended conferences for each researcher. The conferences are ranked in descending order of similarity scores, and the top ***k*** conferences are recommended. The recommendation algorithm is outlined below:

Algorithm 1 Personalized Conference Recommendation Algorithm

Input: Researcher profiles $\{(P_i, R_i)\}_{i=1}^{n}$, Conference data $\{C_j\}_{j=1}^{m}$
Output: Recommended conferences for each researcher
Step 1: Data Collection
for each researcher i **do**
  Create profile (P_i, R_i)
end for
for each conference j **do**
  Extract conference data C_j
end for
Step 2: Feature Extraction
for each researcher i **do**
  Extract themes and topics T_i from (P_i, R_i) using LLM
end for
for each conference j **do**
  Extract themes and topics C_j from conference data using LLM
end for
Step 3: Similarity Calculation
for each researcher i **do**
  for each conference j **do**
    Calculate similarity $\text{sim}(T_i, C_j)$
  end for
end for
Step 4: Recommendation Generation
for each researcher i **do**
  Rank conferences $\{C_j\}$ based on similarity scores $\text{sim}(T_i, C_j)$
  Recommend top k conferences to researcher i
end for

3. IMPLEMENTATION AND OUTCOMES

For our implementation, we employed the powerful language model OpenAI GPT4o [10]. We gave the model a sample dataset for a researcher specialized in AI and cyber security and AI. Various publications of the researcher were provided to the model along with the research interests of researcher. The prompt was as follows:

> ”Please analyze the following texts to extract key themes and topics. For each text, generate an embedding vector that represents its content in a high-dimensional space. Only responde with the results. Here is the publication history of researcher i: N. [Title: ”DATA”, Abstract: ”DATA”]. Here are the research interests of researcher i: N. DATA. For the conference j, here are the attributes: N. [Title: ”DATA”, Abstract: ”DATA”, Keywords: ”DATA”]”

Fig. 2 below shows the analysis of the system along with the recommendation provided.

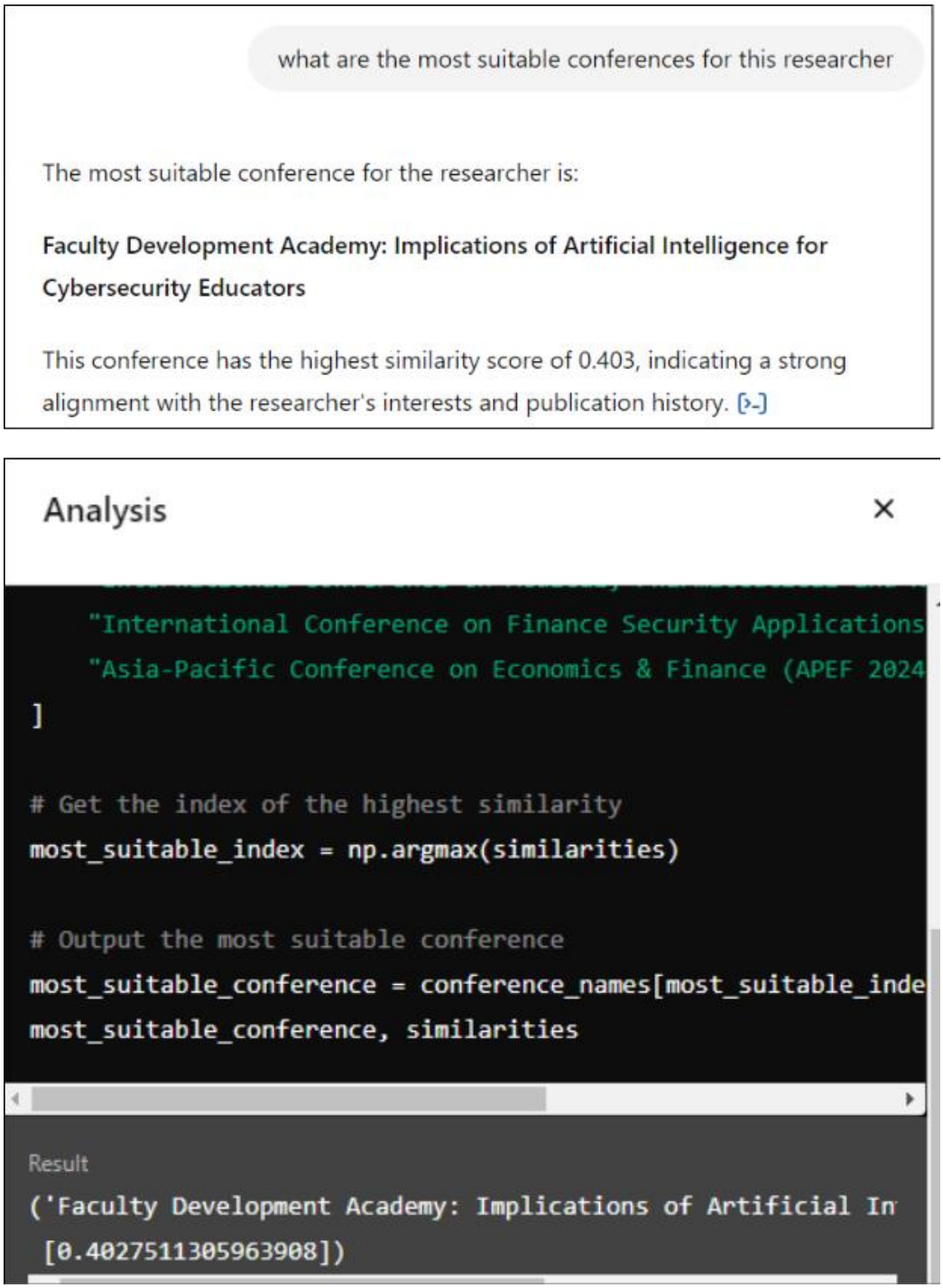

Figure 2: System Response and Recommendations

3.1. Benefits of the Proposed System

The proposed digital database and personalized recommendation system offer several benefits to researchers, students, and professionals in the Arab scientific community. These benefits are summarized in Table I.

Benefits	*Description*
Centralized Information	Provides a unified platform that aggregates comprehensive details about various Arab conferences, including dates, locations, organizers, speakers, and session information
Personalized Recommendations	Utilizes LLMs to analyze researchers' publication histories and interests to offer tailored conference recommendations, ensuring relevance and increasing engagement.
Efficient Conference Discovery	Streamlines the search process for relevant conferences, saving researchers time and effort by quickly identifying conferences that match their specific research areas.
Language Support	Supports multiple languages, particularly Arabic and English, to broaden accessibility and usability for a diverse user base across the Arab world
Enhanced Collaboration	Facilitates communication and collaboration among researchers by providing forums and groups for attendees of the same conferences, fostering academic networking and knowledge sharing
Access to Historical Data	Offers access to information from previous conferences, including papers presented and video recordings, serving as a valuable resource for ongoing research and reference
Improved Academic Engagement	Increases researchers' participation in relevant conferences, enhancing their academic exposure and opportunities for professional development
Up-to-date Information	Ensures that the scientific community stays informed about the latest developments and advancements across various fields, promoting continuous learning and innovation.
Comprehensive Coverage	Includes a wide range of scientific disciplines, ensuring that researchers from diverse fields can find conferences relevant to their interests
User-friendly Interface	Features an intuitive and easy-to-navigate interface, making it simple for users to access and utilize the database effectively.
Real-time Updates	Provides real-time updates on conference schedules, speaker changes, and other relevant information, keeping users informed of the latest developments.
Resource Efficienc	Reduces the need for manual searches and scattered resources, consolidating all necessary information in one place and optimizing researchers' time and efforts
Scalability	Designed to scale with the growing number of conferences and users, ensuring long- term usability and relevance as the scientific community expands.

4. CONCLUSION

The fast pace in which scientific research is conducted, it makes tracking relevant conferences a challenge. In this paper, we present a solution with a holistic digital online database for Arab conferences that incorporates a recommendation system, so as to provide bespoke and powered recommendations by Large Language Models. This system will extract the major themes and topics of researchers' publication histories and interests, detect similarities between such information and conference data, and provide customized recommendations. The system proposed gives efficiency and relevance to the conference discovery, hence higher engagement in the aca- demic community. This is a system which, while offering personalization, language support, and community building, experiences challenges related to data quality, scalability, and privacy that will be addressed in future work. This paper shows that LLMs can be very transformative within the arena of academia, the capabilities of which every researcher will have in order to find relevant conferences and be more academically engaged and collaborative. Future works will focus on enhancing data collection, refining theme extraction, scalability assurance, and integration of user feedback for constant system improvement. This system shall thus be further developed towards bridging the gap between researchers and academic conferences to advance knowledge and innovation.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.

[2] A. Acharya, B. Singh, and N. Onoe, "Llm based generation of item- description for recommendation system," in Proceedings of the 17th ACM Conference on Recommender Systems, 2023, pp. 1204–1207.

[3] R. Y. Maragheh, L. Morishetti, R. Giahi, K. Nag, J. Xu, J. Cho, E. Kor- peoglu, S. Kumar, and K. Achan, "Llm-based aspect augmentations for recommendation systems," 2023.

[4] S. Lubos, T. N. T. Tran, A. Felfernig, S. Polat Erdeniz, and V.-M. Le, "Llm-generated explanations for recommender systems," in Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 2024, pp. 276–285.

[5] A. X. Zhang, A. Bhardwaj, and D. Karger, "Confer: A conference recommendation and meetup tool," in Proceedings of the 19th ACM conference on computer supported cooperative work and social com- puting companion, 2016, pp. 118–121.

[6] S. Li, P. Brusilovsky, S. Su, and X. Cheng, "Conference paper recom- mendation for academic conferences," IEEE Access, vol. 6, pp. 17 153– 17 164, 2018.

[7] H. Jelodar, Y. Wang, G. Xiao, M. Rabbani, R. Zhao, S. Ayobi, P. Hu, and I. Masood, "Recommendation system based on semantic scholar mining and topic modeling on conference publications," Soft Computing, vol. 25, pp. 3675–3696, 2021.

[8] K. Chen, C. B. Choy, M. Savva, A. X. Chang, T. Funkhouser, and S. Savarese, "Text2shape: Generating shapes from natural language by learning joint embeddings," in Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14. Springer, 2019, pp. 100–116.

[9] F. Rahutomo, T. Kitasuka, M. Aritsugi et al., "Semantic cosine similar- ity," in The 7th international student conference on advanced science and technology ICAST, vol. 4, no. 1. University of Seoul South Korea, 2012, p. 1.

[10] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat et al., "Gpt-4 technical report," arXiv preprint arXiv:2303.08774, 2023.