Integration of Large Language Models in Mobile Applications for Statutory Auditing and Finance

Integración de Grandes Modelos de Lenguaje en Aplicaciones Móviles para Revisoría Fiscal y Finanzas

Sergio Robles-Serrano[1], Jesus Rios-Perez[2] y German Sanchez-Torres 3[*]

1. Msc, Universidad del Magdalena, Santa Marta, Colombia, https://orcid.org/0000-0003-4512-0244, sergioroblesas@unimagdalena.edu.co.
2. Ingeniero, Universidad del Magdalena, Santa Marta, Colombia, https://orcid.org/0009-0008-6235-3967, jesusriosdp@unimagdalena.edu.co.
*. PhD, Universidad del Magdalena, Santa Marta, Colombia, https://orcid.org/0000-0002-9069-0732, gsanchez@unimagdalena.edu.co.

Cite this article as: S. Robles-Serrano, J. Rios-Perez y G. Sanchez-Torres. " Integration of Large Language Models in Mobile Applications for Statutory Auditing and Finance", Prospectiva, Vol. 22 N° 1 2024.

Recibido: 07/09/2023 / Aceptado: 14/12/2023
http://doi.org/ 10.15665/rp.v22i1.3334

ABSTRACT

In the current digital age, Artificial Intelligence, with an emphasis on large language models, has gained prominence in various fields such as finance and tax auditing, offering greater efficiency and accuracy in accessing information. This study proposes a software architecture for a mobile application as an intelligent personal assistant in this domain, integrating semantic search and large language models to optimize responses. The methodology included a literature review and a focus on emerging technologies through a technological surveillance study, culminating in an architecture inspired by the Voice Interaction Community Group of the W3C, adapted for non-intent based models with LLM. After developing the application, corporate data was integrated, facilitating semantic searches using a dense passage retrieval scheme and integrating it with language models. The results showed increased efficiency in obtaining financial and tax information and more contextual responses, speeding up data retrieval. This indicates that such integrations can revolutionize how professionals access information. However, it is essential to address ethical, security, and privacy aspects to ensure the reliability and sustained adoption of these tools.

Keywords: Artificial Intelligence; tax auditing; semantic search; Voice Interaction Community Group; large language models.

RESUMEN

En la actual era digital, la Inteligencia Artificial, con énfasis en grandes modelos de lenguaje, ha ganado relevancia en diversas áreas como el campo de las finanzas y revisoría fiscal, ofreciendo mayor eficiencia y precisión en el acceso a la información. Este estudio propone una arquitectura de software para una aplicación móvil como asistente personal inteligente en este dominio, integrando búsqueda semántica y grandes modelos de lenguaje grande para optimizar respuestas. La metodología incluyó revisión literaria y un enfoque en tecnologías emergentes mediante un estudio de vigilancia tecnológica, finalizando en una arquitectura propuesta por el Voice Interaction Community Group de la W3C, adaptada para modelos no basados en intenciones con LLM. Tras desarrollar la aplicación, se integraron datos corporativos, facilitando búsquedas semánticas mediante un esquema de recuperación densa de pasajes e integrándolo con modelos de lenguaje. Los resultados mostraron mayor eficiencia en la obtención de información financiera y fiscal, y respuestas más contextuales, agilizando la recuperación de datos. Esto indica que tales integraciones pueden revolucionar cómo los profesionales acceden a la información. No obstante, es vital abordar aspectos de ética, seguridad y privacidad para asegurar la confiabilidad y adopción sostenida de estas herramientas.

Palabras Clave: Inteligencia Artificial; revisoría fiscal; búsqueda semántica; Voice Interaction Community Group; grandes modelos de lenguaje.

1. Introducción

The financial realm has undergone significant advancements and transformations with the surge of digitalization [1]–[3]. Particularly, voice-based systems and Natural Language Processing (NLP) have begun to play a crucial role in reconfiguring customer service [4] and streamlining operations, with a specific focus on areas such as collections [5]–[8].

Such systems, supported by advanced NLP technologies, such as Amazon Alexa and devices from Google's home ecosystem, have rendered access to banking and financial services ubiquitous and more intuitive [9], [10]. The drive for the adoption of these innovations lies in their potential to revolutionize user experience, with tech companies acting as nexuses between financial institutions and customer-oriented voice platforms [11].

With the recent context of large language models [12], and despite the proliferation of these systems and their increasing adoption, there is a manifest need for research that addresses in detail both their advantages and challenges [13]. Exploring their potential to facilitate a more natural interaction [14] and to enhance inclusivity and inherent challenges, among which are included ethical issues [15], privacy concerns, and associated costs.

In this scenario, our research focuses on formulating a suitable software architecture for the development of a mobile application in the financial and tax audit context. The aim is to integrate advanced language models into a semantic search environment, utilizing dense passage retrieval techniques to provide contextualized answers, all with the objective of facilitating more effective and efficient access to financial information.

Semantic search and Dense Passage Retrieval (DPR) offer significant potential to enhance information retrieval from large data sets [16], [17]. These paradigms go beyond mere word matching and utilize advanced embeddings and machine learning structures to interpret the fundamental intention and semantic features of a query [18]. This results in a retrieval process based on contextual relevance, rather than mere syntactic similarity. DPR also enables the identification of specific segments within extensive documents that align with the information needs of a given query, achieved through dense vectors representing queries and document passages [19]. The combination of these two methodologies results in a faster retrieval process, greater accuracy, and meaningful information, challenging and redefining the paradigms of traditional search algorithms.

To achieve this, we commenced with a thorough review of the literature and technological advancements documented in academic and commercial sources through a Technological Surveillance study [20], [21]. This review granted us a profound understanding of current trends and enabled us to determine the most pertinent methodologies and technologies for our research. Subsequently, we conceived a software architecture that would adapt to the peculiarities of financial and tax data, focusing on its modularity and its capacity to incorporate future enhancements.

This study presents a software architecture for a mobile application in the field of finance and tax auditing. The architecture incorporates semantic search and extensive language models to enhance information retrieval. Researchers conducted a literature review and a technological surveillance investigation to develop the architecture. The application integrates corporate data and enables semantic search through a dense passage retrieval system. The integration of language models significantly improves the retrieval of financial and tax information. The study also highlights the importance of speech technologies and provides a categorization of speech technologies and their main applications. The software architecture is modularized and standardized, allowing for smooth interaction between components and sublayers. The study emphasizes the importance of addressing ethical, security, and privacy concerns for the integration of these capabilities.

2. Methodology

The following describes the methodology used to investigate and evaluate the utilization and incorporation of emerging technologies in the realm of financial services. The methodology is composed of the following stages: technological surveillance, perception studies, software architecture, semantic search, and the integration of LLM (Large Language Models) linguistic models (see Figure 1). Each subsection will present a description of the adopted approach, the tools used, and the fundamental methodological considerations.

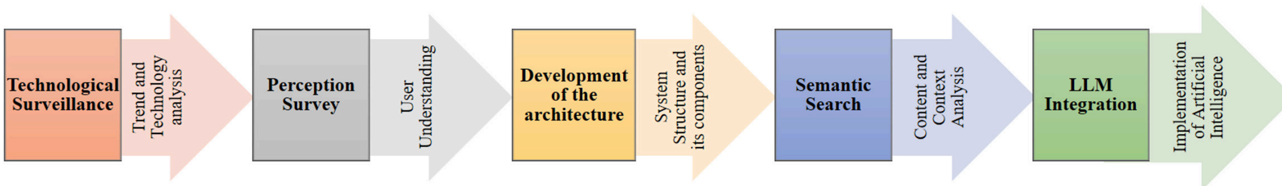

Figure 1. Graphical Overview of the Employed Methodology.
Figura 1. Descripción general gráfica de la metodología usada.

Technological Surveillance

In order to obtain an overview of the trends and practices of the financial services industry, a Technological Surveillance study was implemented. Technological Surveillance is a tool for companies that allows them to acquire external information that can be useful in amplifying their strengths or mitigating their weaknesses concerning technological competencies. This tool has been recognized as a mechanism for analysis and exploration to identify future technologies beneficial to organizations. Over time, the definitions of Technological Surveillance have expanded, integrating a more comprehensive perspective of Competitive Intelligence, yet still grounded in the original foundation of surveillance [22]. Various methodologies have been suggested in the literature, converging on similar phases such as the search and analysis of information. However, discrepancies exist in the conceptualization of certain stages, such as whether decision-making based on the collected information should be an intrinsic part of the process or simply result from it.

In the context of this work, a specific methodology of technological surveillance was adopted, composed of five crucial phases: definition and delimitation of key elements, identification and selection of information sources, search and data collection, filtering and analysis, and construction of reports (see Figure 2).

Figure 2. General Steps of the Technological Surveillance Study Development.
Figura 2. Pasos generales del desarrollo del estudio de Vigilancia Tecnológica.

It is essential to note that, under this methodology, decision-making is considered a byproduct of the Technological Surveillance process and not a phase within it. This perspective aligns with previous academic works [21], [23], [24]. The adopted phases range from defining thematic areas of interest to the detailed presentation of trends and technological predictions relevant to the organization.

Perception Studies Methodology

The methodology used for the perception study aimed primarily to understand the perceptions, preferences, and needs of customers and workers concerning the use of voice assistants and the current functionalities of the platform.

The incorporation and massification of Intelligent Personal Assistants (IPAs) have sparked debates about their efficacy and the protection of personal information. Current research [25]–[27]but a key problem with speech as an interaction modality is how to scaffold accurate mental models of voice assistants, a task complicated by privacy and security concerns. We present the results of a survey of voice assistant users (n=1314 indicates that, although the current version of IPAs exhibits the ability to continuously monitor and actively offer services, apprehensions regarding privacy and security continue to prevail. User expectations span a wide range of services; however, their inclination to reveal data depends on several factors, including the sensitivity of the information and their level of familiarity with the device.

The simultaneous challenge underscores the need to address perceptions about how users construct mental models of how these assistants operate [28], [29], as well as concerns about privacy. The amplification of concerns about trust and privacy is reinforced by specific explanations, indicating the need to conduct perception studies to ensure the appropriate implementation and design of voice assistants in corporate environments [15], [30], [31]it has gained a huge popularity amongst research as well as industrial community. Recently, many studies have been published to show the effectiveness, efficiency, integration, and sentiments of chatGPT and other LLMs. In contrast, this study focuses on the important aspects that are mostly overlooked, i.e. sustainability, privacy, digital divide, and ethics and suggests that not only chatGPT but every subsequent entry in the category of conversational bots should undergo Sustainability, PrivAcy, Digital divide, and Ethics (SPADE.

For this, an instrument was structured, divided into three key sections: the voice assistant, the current platform, and the project framework. The sample chosen for this study consisted of a group of company customers and workers, selected with the aim of capturing a representative perception. Regarding the data collection procedure, participants were invited to participate voluntarily. The questionnaire was administered to them, taking measures to ensure that all questions were adequately understood. Subsequently, the collected data were analyzed quantitatively.

Software Architecture Development

The reference software architecture chosen for this project is the proposed standard for intelligent personal assistants developed by the Voice Interaction Community Group [32]interaction with voice applications has become much more flexible, with a user-initiated dialog style and significantly fewer constraints on spoken input. \n\nMany of these new applications take the form of \"virtual assistants\". These include general-purpose assistants (for example, Siri, Cortana, Google Now and Alexa of the World Wide Web Consortium (W3C) [33]. The proposed architecture for the voice interface is based on a client-server-Service Provider model [34]. In this scheme, the client, symbolized by the personal audio device, performs the function of acquiring and playing back audio from the user. The server, playing the central role in the backend, is responsible for processing voice commands and generating appropriate responses. Using a natural language processing system, the received commands are decoded, and the Service Provider is requested for corporate information semantically related to the intent of the client's request (see Figure 3).

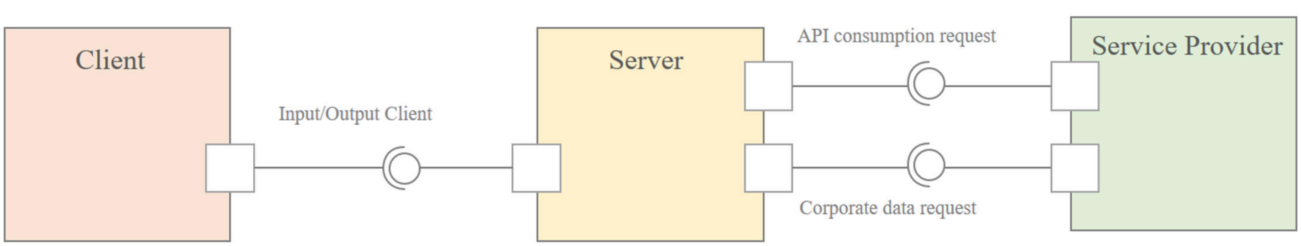

Figure 3. General Software Architecture of Client-Server-Service Provider.
Figura 3. Arquitectura de software general Cliente-Servidor-Proveedor de Servicios.

The flow of voice interaction is structured through a series of consecutive steps that ensure smooth communication between the user and the device. Firstly, the user issues a verbal instruction that is recorded by the client. Once the text is converted and sent to the server, the server interprets it through natural language processing. It performs the retrieval of corporate data semantically associated with the user's request through semantic search. The server has the ability to access various data sources through the Service Provider's Application Programming Interface (API). Once the response has been determined, it is synthesized and sent to the client, who plays it for the end user. The architecture allows for versatility by enabling adjustments according to the particularities of the device or application under consideration.

Semantic Search and Integration of Large Language Model

Large Language Models (LLMs) have demonstrated exceptional performance across various natural language processing tasks. However, as they grow in size, they face significant challenges in terms of computational costs. Moreover, they often lack efficient, domain-specific understanding [35] or up-to-date information [36], [37]. Nevertheless, their performance may be lower in specific domain tasks that require specialized knowledge [37], [38], in specialized fields such as tax auditing and finance. Thus, they are usually supported by data augmentation methods derived from the user's query. The Dense Information Retrieval (DIR) model functions as a knowledge repository [18], [19], addressing limitations related to coherence and long-term memory observed in LLM systems. The synergy between knowledge and reasoning enhances the effectiveness of an information service system by increasing its level of intelligence and reliability.

The dense passage retrieval method emerges as an alternative to traditional information retrieval techniques, which tend to depend on sparse features. This system generates low-dimensionality representations for both queries and documents. The fundamental stages of the Dense Passage Retriever (DPR) include encoding, mapping, query processing, vector comparison, ranking, and retrieval. Encoding is a process by which textual content is transformed into vectors of reduced dimensions that are located in a shared embedding space. The mapping process assigns each document and query a vector in the corresponding space. Queries are transformed into dense vector representations [39], [40] and subsequently compared with vectors corresponding to documents stored in the database. Finally, documents are categorized based on their degree of similarity to the query vector.

The emergence of large language models has played a significant role as a determining factor in the revolution of semantic search. Models like GPT-3 and its subsequent iterations employ deep neural network structures that have been trained on extensive datasets to achieve a sophisticated understanding of textual content at a semantic level. These models possess a vast number of adjustable parameters, allowing them to produce coherent responses, complete text sections, and even draft essays on designated topics with a notable level of fluency and accuracy.

Incorporating these expansive language models into semantic search systems presents several notable benefits: they have the ability to discern variations in queries that are similar but differ in intent. This ability allows them to provide more personalized results according to the user's specific requirements. They generate dynamic responses, meaning they have the ability to generate responses that are not limited to retrieving information but are modulated with the information with which they have been trained. This capability enables the provision of more comprehensive and personalized solutions. They allow natural interaction through natural language understanding. Additionally, these models have the ability to adapt and improve their performance over time, particularly when subjected to regular retraining using new data.

3. Results and Discussion

Technological Surveillance

Upon conducting Technological Surveillance, an increase in research interest regarding the development of voice-based technologies was evidenced. This search was performed in two databases (Web of Science and Scopus) between the years 2011 and 2021. A total of 7975 scientific publications were found in the area of study. The analysis of results shows a growing trend of publications on the topic throughout the last decade. The behavior of the number of publications shows an increase of more than double the number of publications in the last five years and an increase of more than 400% since 2011. This behavior is illustrated in Figure 4.

On the other hand, the highest number of publications collected from the databases originates in the United States (20%), India (15%), and China (14%), which present approximately half of the total results found (approx. 49%). They are followed by Germany (6%), the United Kingdom (6%), Japan (4%), France (4%), Italy (3%), Spain (3%), and Canada (3%). These countries are the 11 countries that report the highest number of publications in the scientific databases of the defined theme.

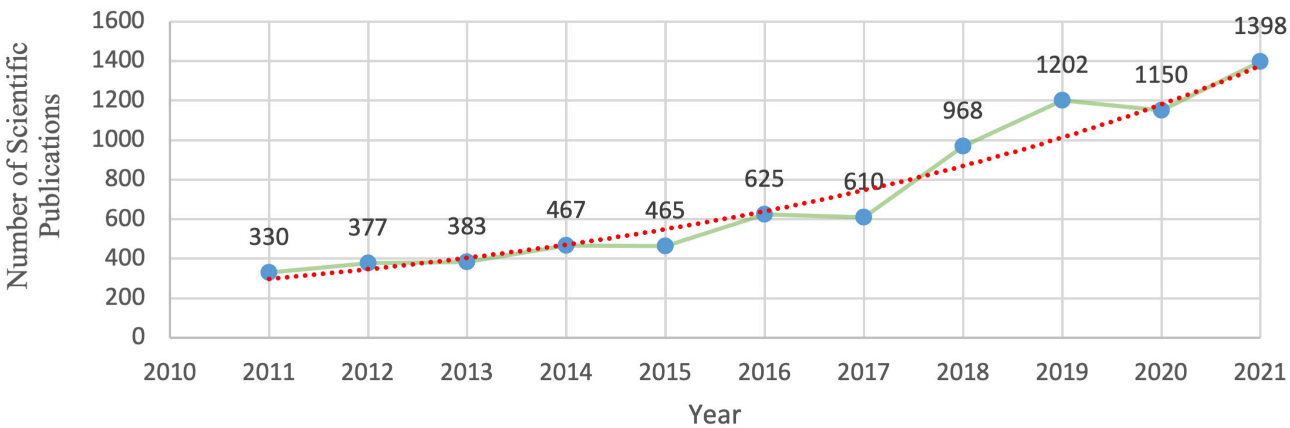

Figure 4. Behavior of Scientific Publications in the Area of Voice-Based Systems.
Figura 4. Comportamiento de las publicaciones científicas en el área de sistemas basados en voz.

In the distribution of publications according to the type of scientific document compiled in various databases, Conference Publications predominate with 75%, followed by Scientific Articles which constitute 22%, while Book Chapters represent only 3%. This pattern highlights a marked academic interest in research related to voice-based technologies.

A general taxonomy has been constructed from the analysis of literature related to speech technologies, which has three main classifications, as shown in Figure 5: Speaker Recognition (SR), Speech Recognition (SPR), and Voice Generation (TTS).

According to [41]there is a need for voice bio-metrics, through which we can validate a person's identity over a wireless medium like mobile. Voice bio-metrics have many advantages over other bio-metrics like retinal scans and fingerprints because in the case of voice bio-metrics there is no need for physical presence. Speaker recognition (SR, speaker recognition refers to the procedure of discerning and verifying an individual's identity by analyzing their vocal attributes. The distinctive character of this phenomenon arises from a combination of inherent elements, ranging from anatomical characteristics to linguistic variations. It is worth mentioning that specific attributes, such as voice pitch, exhibit consistency, allowing their use as a distinctive feature.

Regarding speech recognition, in [42], it is described that the scope of voice recognition is limited to converting auditory expressions into written forms using alphabetical symbols. The convergence of acoustics and linguistics, facilitated by advances in artificial intelligence and signal processing, is evident in a wide range of applications, ranging from dictation to automatic translation. The categorization and subcategorization of this technology are determined by various attributes that include vocabulary size, user recognition capabilities, and conversion methods.

In speech generation, in [43], the process of converting written text into audible voice, known as voice generation or text-to-speech (TTS), is described. The scope of its applicability ranges from automated call centers to digital literature.

Figure 6 shows the main application areas for each category.

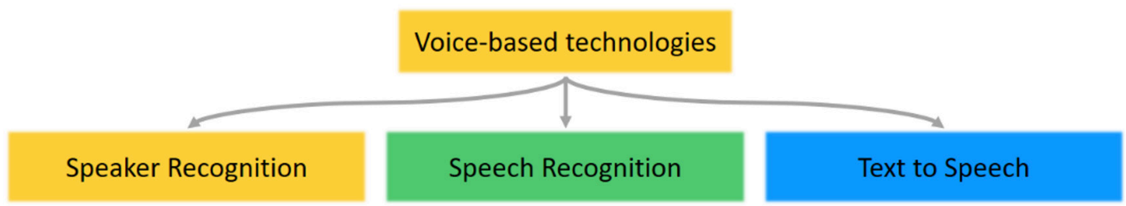

Figure 5. Taxonomy of Voice-Based Applications.
Figura 5. Taxonomía de las aplicaciones basadas en voz.

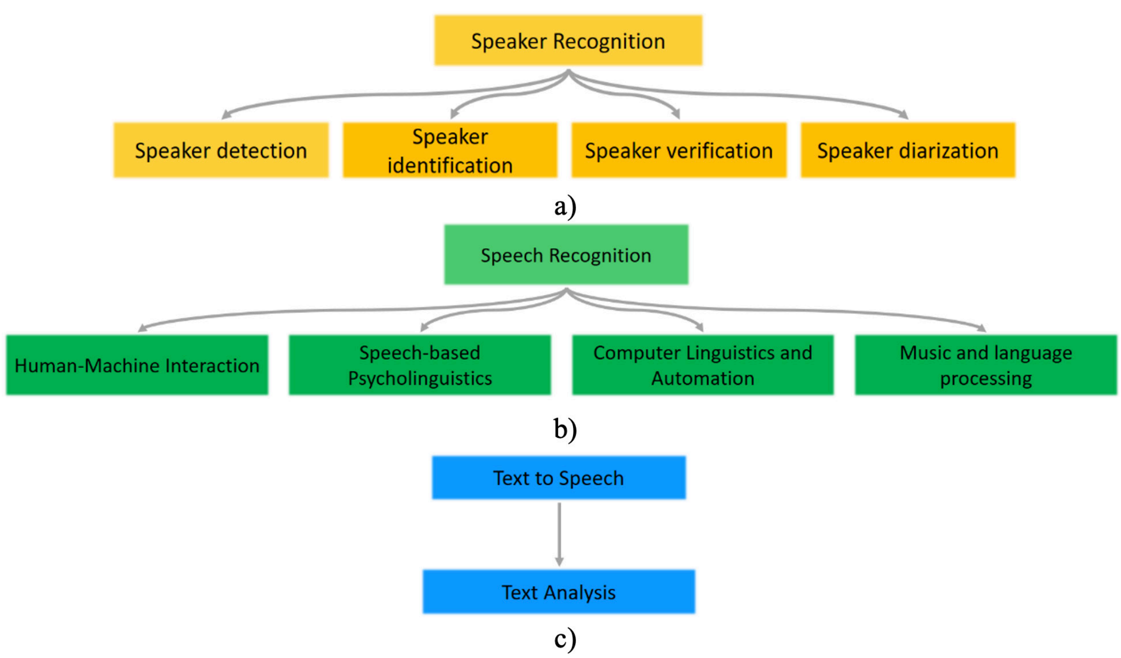

Figure 6. Primary Applications in Each Category of Voice-Based Systems.
Figura 6. Principales aplicaciones en cada categoría de sistemas basados en voz.

Software architecture development

The architecture is composed of three components (see Figure 7):

1. Client: This layer has two components that interact directly with the user: microphone (receives the speaker's voice) and speakers (returns a coherent response to the speaker's request). Local services perform functionalities such as launching the application.
2. Server: It has several sub-layers which are:

- IPA Service: This layer receives the audio and simultaneously sends it to the local IPA (the recipient is configured based on the amount of information received in the audio, referring to metadata).
- ASR: The Automated Speech Recognizer, receives an audio stream and generates a recognition hypothesis as text strings. Optionally, it can generate multiple recognition hypotheses. It allows updating the history with determined recognition hypotheses.
- Intent Core: Extracts the meaning of intentions (group of expressions with similar meaning) and associated entities (captures additional information from an intention, such as dates) from an expression as a text string.
- Dialog Manager: Receives semantic information determined from the user's input, updates the dialog history, its internal state, decides the next steps to continue a dialog, and provides results, mainly as synthesized or recorded statements. Conceptually, the dialog manager defines the playground used by dialog boxes and significantly contributes to the user experience.
- Context: Contains short and long-term information to maintain a logical conversation with the user, and helps guide the conversation.

3. Service Provider:

- Service Provider Selection: Provides access to all data providers, external services, and known IPA providers. Assigns IPA provider sets to the intent sets of the dialog layer.
- Data Provider: Obtains data from various sources for use in interaction.
- External Services: Provide data from sources external to the system, for example, data obtained from third-party servers.
- Service Providers (API): Provide IPAs that can interact with users in the application.

The flow of request input from the user begins when interacting with the application through the microphone, and the client layer makes a request to provide all local information and return it to the client layer to then make use of the IPA services (to be able to handle the information as appropriate), the audio is transformed into text, consequently, natural language processing is done, this interacts with the context layer to carry on a logical conversation. Finally, it interacts with the API or data layer, it goes through the provider registration layer and then the data authentication is done, so that later the IPA provider does the text recognition.

The response flow of the architecture begins with the service provider that gives the response to the service provider selection layer. Then it reaches a dialog management layer, this communicates with the context layer to carry on a more coherent conversation, then it interacts with the dialog layer. As one of the last steps, the text is passed to audio and the response is returned to the client layer, finally, the personal assistant's response will be played through our user's speakers.

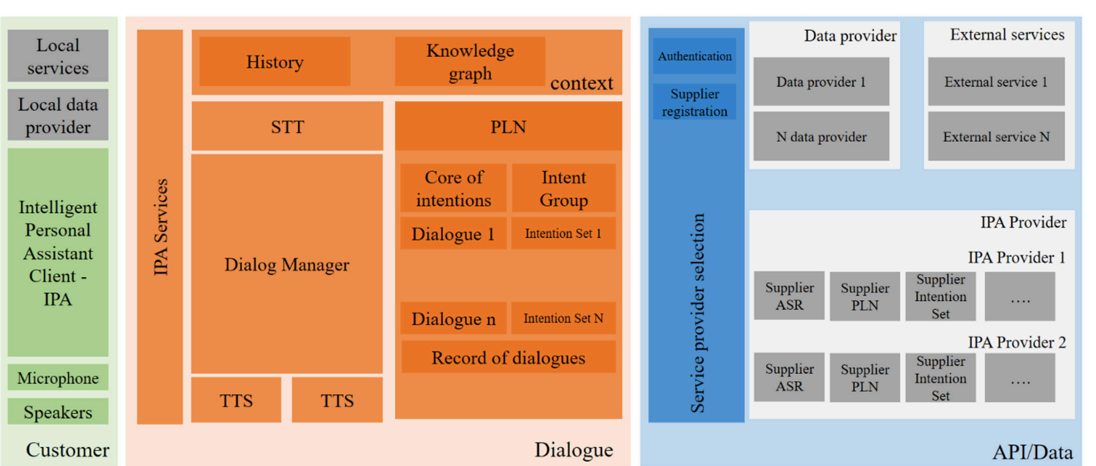

Figure 7. Software Architecture Used, Modified from [33].
Figura 7. Arquitectura de software utilizada, modificada de [33].

Semantic Search and Integration with Large Language Models

To achieve precise and efficient results, specialized systems require substantial amounts of data, as well as a deep understanding of semantics. With the growing use of large language models (LLMs) in corporate environments, there is a need for a framework that enhances the semantic content of stored data, particularly in relation to question and answer (QA) systems. This enrichment aims to meet the information requirements of LLMs and improve the accuracy of semantic searches, which are crucial for retrieving information in various contexts. Figure 8 shows the scheme built for semantic search support and integration with large language models.

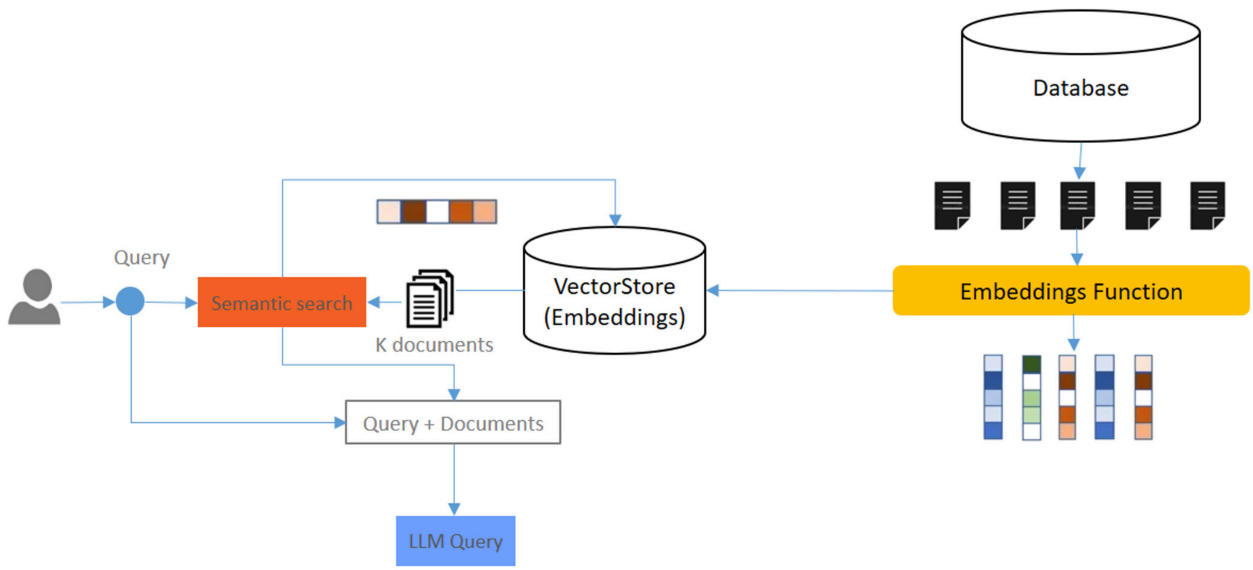

Figure 8. Semantic Search Scheme and Use of a LLM.
Figura 8. Esquema de búsqueda semántica y uso de LLM.

Qualitative Evaluation of the System

The use of qualitative evaluation is an essential methodology for assessing the efficacy and accuracy of systems or models. The following subsection provides an analysis of the performance of semantic search and language model integration. A variety of evaluation metrics were established, as illustrated in Table 1, spanning a spectrum of factors including usability and customer service excellence. These metrics were evaluated using a numerical scale ranging from 1 to 10. This evaluation was developed under the criterion of three tax audit experts from the company. The profiles of the experts are presented in Table 2. These experts interacted with the system within specific scenarios that were carefully outlined according to a predetermined set of intentions. The aim is to evaluate the accuracy and efficacy of the system in recognizing and addressing these intentions. The robustness of the evaluation establishes a solid foundation for future analyses of the acquired results. Figure 9 describes the results of this evaluation.

Each of the aspects and their valuation scale are specified below, in Table 1. Table 2 presents the selected expert profiles.

The experts were also required to evaluate aspects related to the application. The criteria and their results are shown in Table 3.

The results of the qualitative tests indicate a positive reception from the experts. Overall, the application was weighed in terms of ease of use, interface clarity, speed, and responsiveness, with consistently high scores in all these areas. The level of satisfaction, content quality, and customer service were also highlighted with favorable ratings. Together, the qualitative evaluation shows that the application is efficient, easy to use, and satisfactory for users.

Table 1. Metrics for the evaluation of inference capacity and their rating scale.

Tabla 1. Métricas para la evaluación de la capacidad de inferencia y su escala de valoración.

Evaluation criterion	Rating scale
Ease of Use	1-10
Interface Clarity	1-10
Response Speed	1-10
Level of Satisfaction	1-10
Responsiveness	1-10
Content Quality	1-10
Customer Service Quality	1-10

Table 2. Description of the Evaluator Experts.

Tabla 2. Descripción de los expertos evaluadores.

Expert	Description
Expert 1	Male, 35 years old, native Spanish speaker with a Latin American accent, specialist in user experience (UX).
Expert 2	Female, 30 years old, native Spanish speaker with a Latin American accent, professional in international trade.
Expert 3	Male, 42 years old, native Spanish speaker with a Latin American accent, software engineer with experience in voice application development.

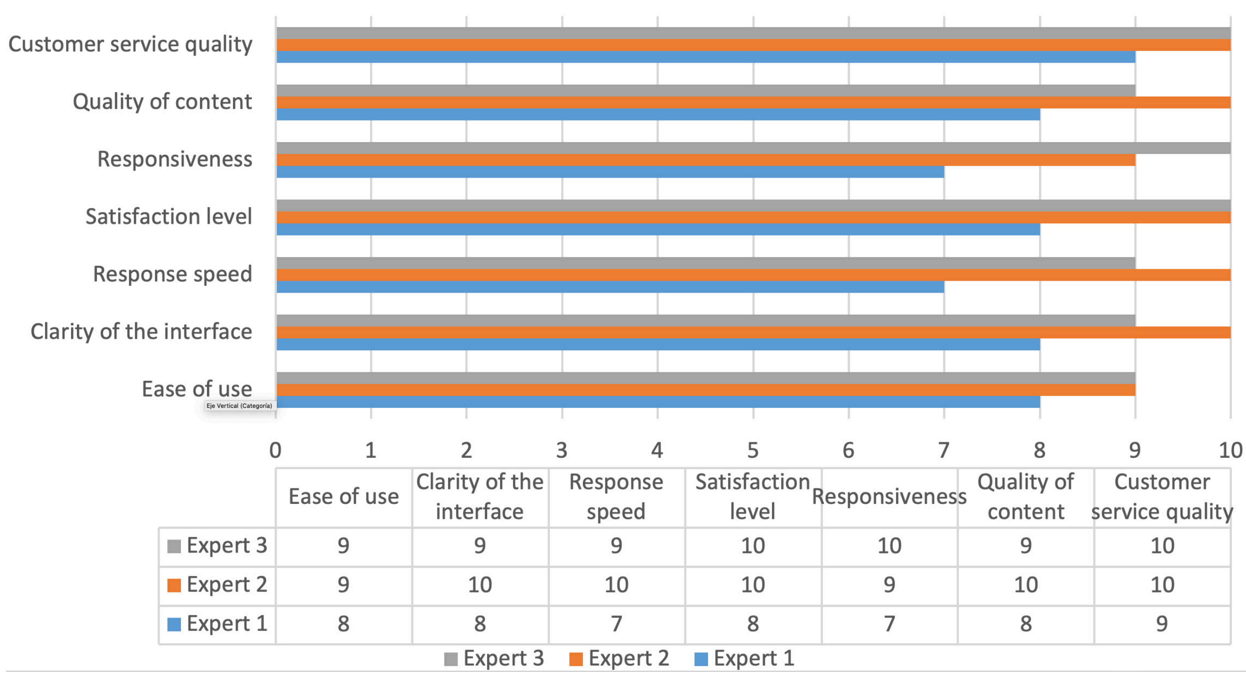

Figure 9. Evaluation Results of Metrics with Potential Users.

Figure 9. Resultados de evaluación de métricas con usuarios potenciales.

Table 3. Evaluation Results of Metrics related with application experience.
Tabla 3. Resultados de evaluación de métricas relacionadas con la aplicación.

Evaluation criterion	Expert 1	Expert 2	Expert 3
Ease of Use	8	9	9
Interface Clarity	8	10	9
Response Speed	7	10	9
Level of Satisfaction	8	10	10
Responsiveness	7	9	10
Content Quality	8	10	9
Customer Service Quality	9	10	10

4. Conclusions

Technological Surveillance reveals a significant increase in the scientific community's attention towards the advancement of voice-based technologies over the last decade. The field of research has experienced a substantial increase in publications, as indicated by a notable growth of more than 400% since 2011, demonstrating its rising importance. Conference presentations and scientific articles constitute the main forms of academic contributions, highlighting a dynamic and active exchange of knowledge within the academic community. Analyzing the publications from a geographical perspective reveals that countries like the United States, India, and China exhibit prominent leadership in terms of their contributions, and together, represent almost half of the overall results. These technologically advanced entities have demonstrated their prominence as key centers for voice technology research. Research in this area is primarily dominated by the disciplines of Computer Science, Engineering, and Mathematics. However, it is noteworthy that there is significant participation from fields such as Social Sciences, and Arts and Humanities. The aforementioned statement highlights the interdisciplinary nature of voice technologies and their impact on various sectors of society.

The utilized software architecture allowed for modular and standardized development. In which there is a collaboration of various components and sub-layers, which collectively enable tasks such as voice input and the delivery of consistent answers.

The integration of a large language model has resulted in greater efficiency in semantic search and information retrieval. The integration of these models with information retrieval models has played a crucial role in addressing their limitations and enhancing the precision and contextual relevance of textual content processing.

Dense retrieval has the potential to address the limitations of conventional retrieval methods by offering concise representations of queries and documents. The use of dense vectors to convert textual content promises to enhance the precision and relevance of future retrieval systems.

References

[1] B. Huang, «On the Practical Application of Computer Technology in Finance and Tax Audit», J. Phys.: Conf. Ser., vol. 1744, n.o 3, p. 032042, feb. 2021, doi: 10.1088/1742-6596/1744/3/032042.

[2] G. S. Jayesh, D. Novaliendry, S. K. Gupta, A. K. Sharma, y B. Hazela, «A Comprehensive Analysis of Technologies for Accounting and Finance in Manufacturing Firms», ECS Trans., vol. 107, n.o 1, p. 2715, abr. 2022, doi: 10.1149/10701.2715ecst.

[3] B. Adiloglu y N. Gungor, «THE IMPACT OF DIGITALIZATION ON THE AUDIT PROFESSION: A REVIEW OF TURKISH INDEPENDENT AUDIT FIRMS», JBEF, vol. 8, n.o 4, Art. n.o 4, dic. 2019.

[4] S. Patel, Y.-T. Chiu, M. S. Khan, J.-G. Bernard, y T. A. T. Ekandjo, «Conversational Agents in Organisations: Strategic Applications and Implementation Considerations», JGIM, vol. 29, n.o 6, pp. 1-25, nov. 2021, doi: 10.4018/JGIM.20211101.oa53.

[5] L. Nuñez Aguilar, «SISTEMA IVR PARA LA MEJORA DE LA GESTION DE COBRANZA DE LA EMPRESA CONSORCIO DE TECNOLOGÍA E INNOVACIÓN S.A.C., JAÉN 2017», Repositorio Institucional - USS, 2018, Accedido: 3 de septiembre de 2023. [En línea]. Disponible en: http://repositorio.uss.edu.pe//handle/20.500.12802/5081

[6] F. Huseynov, «Chatbots in Digital Marketing: Enhanced Customer Experience and Reduced Customer Service Costs», en Contemporary Approaches of Digital Marketing and the Role of Machine Intelligence, IGI Global, 2023, pp. 46-72. doi: 10.4018/978-1-6684-7735-9.ch003.

[7] K. S. Carranza Rodríguez y G. M. Carranza rodríguez, «Sistema de Información para el proceso de Gestión de Cobranzas de carteras morosas en la empresa Crédito y Cobranzas SAC. ChiclayoLambayeque», may 2018, Accedido: 3 de septiembre de 2023. [En línea]. Disponible en: http://repositorio.unprg.edu.pe/handle/20.500.12893/1864

[8] Y. Y. Lopez Vitor y R. C. Rojas Hilario, «Asistente virtual para el seguimiento de cobranza en una empresa de envases metálicos usando lenguaje natural», Universidad Peruana de Ciencias Aplicadas (UPC), sep. 2021, Accedido: 3 de septiembre de 2023. [En línea]. Disponible en: https://repositorioacademico.upc.edu.pe/handle/10757/658517

[9] H. Lee, Voice User Interface Projects: Build voice-enabled applications using Dialogflow for Google Home and Alexa Skills Kit for Amazon Echo. Packt Publishing Ltd, 2018.

[10] M. J. Sánchez-Franco, F. J. Arenas-Márquez, y M. Alonso-Dos-Santos, «Using structural topic modelling to predict users' sentiment towards intelligent personal agents. An application for Amazon's echo and Google Home», Journal of Retailing and Consumer Services, vol. 63, p. 102658, nov. 2021, doi: 10.1016/j.jretconser.2021.102658.

[11] T. Lau y B. Leimer, «The era of connectedness: How AI will help deliver the future of banking», Journal of Digital Banking, vol. 3, n.o 3, pp. 215-231, ene. 2019.

[12] A. Anand, V. V, A. Anand, y V. Setty, «Query Understanding in the Age of Large Language Models». arXiv, 28 de junio de 2023. doi: 10.48550/arXiv.2306.16004.

[13] P. D. 29 N. 2019 |Updated: 29-N.-2019 |Category: O. |Author: W. |Member L. Gold |Points: 0 |, «The 12 Advantages and Disadvantages of Voice User Interface», Techulator. Accedido: 29 de noviembre de 2022. [En línea]. Disponible en: https://www.techulator.com/resources/18784-the-12-advantages-and-disadvantages-of-voice-user-interface

[14] J. Kiseleva, K. Williams, A. Hassan Awadallah, A. C. Crook, I. Zitouni, y T. Anastasakos, «Predicting User Satisfaction with Intelligent Assistants», en Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, en SIGIR '16. New York, NY, USA: Association for Computing Machinery, jul. 2016, pp. 45-54. doi: 10.1145/2911451.2911521.

[15] S. A. Khowaja, P. Khuwaja, y K. Dev, «ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review». arXiv, 13 de abril de 2023. doi: 10.48550/arXiv.2305.03123.

[16] V. Karpukhin et al., «Dense Passage Retrieval for Open-Domain Question Answering». arXiv, 30 de septiembre de 2020. doi: 10.48550/arXiv.2004.04906.

[17] D. S. Sachan et al., «Improving Passage Retrieval with Zero-Shot Question Generation». arXiv, 2 de abril de 2023. doi: 10.48550/arXiv.2204.07496.

[18] X. Ma, M. Li, K. Sun, J. Xin, y J. Lin, «Simple and Effective Unsupervised Redundancy Elimination to Compress Dense Vectors for Passage Retrieval», en Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, nov. 2021, pp. 2854-2859. doi: 10.18653/v1/2021.emnlp-main.227.

[19] Y. Wang, H. Ma, y D. Z. Wang, «LIDER: An Efficient High-dimensional Learned Index for Large-scale Dense Passage Retrieval». arXiv, 9 de octubre de 2022. doi: 10.48550/arXiv.2205.00970.

[20] F. O. Cruz?Páez y O. Vanegas?Flórez, «Vigilancia tecnológica, inteligencia competitiva y cultura organizacional universidad de Cundinamarca Facatativá [1]», Política, Globalidad y Ciudadanía, vol. 6, n.o 12, pp. 84-101, 2020.

[21] B. Fernandez, S. Pérez, y F. Del-Valle-Gastaminza, «Metodología para la implantación de sistemas de vigilancia tecnológica y documental: El caso del proyecto INREDIS», Investigación bibliotecológica, vol. 23, pp. 149-177, dic. 2009, doi: 10.22201/iibi.0187358xp.2009.49.21393.

[22] P. M. Ortega, «Vigilancia e inteligencia competitiva: fundamentos e implicaciones», Revista madri+d. Monografía: revista de investigación en gestión de la innovación y tecnología, n.o 7 (Agosto), pp. 15-22, 2003.

[23] A. Nosella, G. Petroni, y R. Salandra, «Technological change and technology monitoring process: Evidence from four Italian case studies», Journal of Engineering and Technology Management, vol. 25, pp. 321-337, dic. 2008, doi: 10.1016/j.jengtecman.2008.10.001.

[24] L. Rey, «Informe APEI sobre vigilancia tecnológica», Informes APEI, No. 4, 2009, ISBN 978-84-692-7999-1, dic. 2009.

[25] W. Seymour y J. Such, «Ignorance is Bliss? The Effect of Explanations on Perceptions of Voice Assistants», Proc. ACM Hum.-Comput. Interact., vol. 7, n.o CSCW1, p. 64:1-64:24, abr. 2023, doi: 10.1145/3579497.

[26] M. Tabassum et al., «Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants», Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 3, n.o 4, p. 153:1-153:23, sep. 2020, doi: 10.1145/3369807.

[27] L. Mirghaderi, M. Sziron, y E. Hildt, «Investigating user perceptions of commercial virtual assistants: A qualitative study», Frontiers in Psychology, vol. 13, 2022, Accedido: 3 de septiembre de 2023. [En línea]. Disponible en: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.944714

[28] A. Mishra, A. Shukla, y S. K. Sharma, «Psychological determinants of users' adoption and word-of-mouth recommendations of smart voice assistants», International Journal of Information Management, vol. 67, p. 102413, dic. 2022, doi: 10.1016/j.ijinfomgt.2021.102413.

[29] S. K. Sharma, R. De, A. Jeyaraj, y R. Raman, «Guest editorial: Re-imagining diffusion and adoption of emerging technologies», International Journal of Information Management, vol. 67, p. 102541, dic. 2022, doi: 10.1016/j.ijinfomgt.2022.102541.

[30] A. Farooq, D. Jeske, P. van Schaik, y M. Moran, «Voice Assistants: (Physical) Device Use Perceptions, Acceptance, and Privacy Concerns», en The Role of Digital Technologies in Shaping the Post-Pandemic World, S. Papagiannidis, E. Alamanos, S. Gupta, Y. K. Dwivedi, M. Mäntymäki, y I. O. Pappas, Eds., en Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 485-498. doi: 10.1007/978-3-031-15342-6_37.

[31] D. Pal, M. D. Babakerkhell, y X. Zhang, «Exploring the Determinants of Users' Continuance Usage Intention of Smart Voice Assistants», IEEE Access, vol. 9, pp. 162259-162275, 2021, doi: 10.1109/ACCESS.2021.3132399.

[32] «Voice Interaction Community Group». Accedido: 2 de septiembre de 2023. [En línea]. Disponible en: https://www.w3.org/community/voiceinteraction/

[33] «W3C», W3C. Accedido: 3 de septiembre de 2023. [En línea]. Disponible en: https://www.w3.org/

[34] «Intelligent Personal Assistant Interfaces». Accedido: 3 de septiembre de 2023. [En línea]. Disponible en: https://w3c.github.io/voiceinteraction/voice%20interaction%20drafts/paInterfaces/paInterfaces.htm

[35] A. Agarwal, S. Gawade, A. P. Azad, y P. Bhattacharyya, «KITLM: Domain-Specific Knowledge InTegration into Language Models for Question Answering». arXiv, 7 de agosto de 2023. doi: 10.48550/arXiv.2308.03638.

[36] A. Halevy y J. Dwivedi-Yu, «Learnings from Data Integration for Augmented Language Models». arXiv, 10 de abril de 2023. doi: 10.48550/arXiv.2304.04576.

[37] Z. Luo et al., «Augmented Large Language Models with Parametric Knowledge Guiding». arXiv, 18 de mayo de 2023. doi: 10.48550/arXiv.2305.04757.

[38] M. T. R. Laskar, M. S. Bari, M. Rahman, M. A. H. Bhuiyan, S. Joty, y J. X. Huang, «A Systematic Study and Comprehensive Evaluation of ChatGPT on Benchmark Datasets». arXiv, 5 de julio de 2023. doi: 10.48550/arXiv.2305.18486.

[39] F. Lashkari, E. Bagheri, y A. A. Ghorbani, «Neural embedding-based indices for semantic search», Information Processing & Management, vol. 56, n.o 3, pp. 733-755, may 2019, doi: 10.1016/j.ipm.2018.10.015.

[40] H. Zamani y W. B. Croft, «Estimating Embedding Vectors for Queries», en Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval, en ICTIR ’16. New York, NY, USA: Association for Computing Machinery, sep. 2016, pp. 123-132. doi: 10.1145/2970398.2970403.

[41] K. N. R. K. R. Alluri y A. K. Vuppala, «Chapter 7 - A study on the emotional state of a speaker in voice bio-metrics», en Advances in Ubiquitous Computing, A. Neustein, Ed., en Advances in ubiquitous sensing applications for healthcare. , Academic Press, 2020, pp. 223-236. doi: 10.1016/B978-0-12-816801-1.00007-4.

[42] H. Liu, «Chapter 1 - Introduction», en Robot Systems for Rail Transit Applications, H. Liu, Ed., Elsevier, 2020, pp. 1-36. doi: 10.1016/B978-0-12-822968-2.00001-2.

[43] J. P. Campbell, W. M. Campbell, A. V. McCree, C. J. Weinstein, y S. M. Lewandowski, «Chapter 10 - Cognitive Services for the User», en Cognitive Radio Technology (Second Edition), B. A. Fette, Ed., Oxford: Academic Press, 2009, pp. 305-324. doi: 10.1016/B978-0-12-374535-4.00010-2.