Creating Digital Identity for Arabic National Libraries from Content Analysis of Missions & Vision Statements Using Machine Learning Algorithms. A Practical Study

د. مؤمن سيد عثمان يس النشرتي

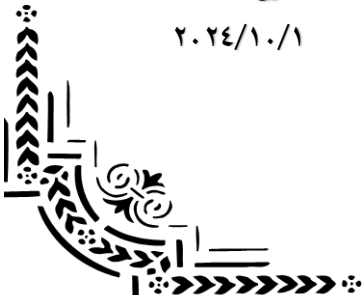### مدرس بقسم المكتبات والوثائق والمعلومات كلية الآداب جامعة القاهرة

Abstract:

This study explores the utilization of ML content analysis algorithms to create a robust digital identity for Arabic national libraries by analyzing their mission and vision statements. The proposed approach aims to align the digital identity with the core roles, values, tasks, and strategic goals of these libraries, thereby fostering a coherent and impactful online existing. the study employed a practical method for applying machine learning model on deriving digital identities depending on the libraries mission statement and vision. the population of study consisted of a selection of six Arabic national libraries, where these libraries provide access to their mission and vision statements by their websites. The most importance findings were that Topic modeling with LDA algorithm is a powerful methodology for discovering hidden themes in Arabic National Libraries vision and mission statements to build a digital identity for these libraries[1].

المستخلص

هدفت هذه الدراسة إلى توظيف خوارزميات تحليل محتوى التعلم الآلي لإنشاء هوية رقمية قوية للمكتبات الوطنية العربية من خلال تحليل بيانات كل مهمتها ورؤيتها. كما تهدف الدراسة إلى تقديم نهج مقترح لمواءمة الهوية الرقمية مع الأدوار والقيم والمهام والأهداف الإستراتيجية لهذه المكتبات، وبالتالي تعزيز وجودها وجعله متماسك ومؤثر عبر الإنترنت، وقد اعتمدت الدراسة في ذلك على المنهج التجريبي لتطبيق نموذج التعلم الآلي في استخلاص الهويات الرقمية اعتمادًا على بيان مهمة المكتبة ورؤيتها، وقد تكوّن مجتمع الدراسة من مجموعة مختارة من ست مكتبات وطنية عربية، حيث توفر هذه المكتبات إمكانية الوصول إلى بيانات رسالتها ورؤيتها من خلال مواقعها الإلكترونية، وكانت أهم النتائج لهذه الدراسة أن نمذجة المواضيع باستخدام خوارزمية LDA هي منهجية قوية لاكتشاف المواضيع الخفية في رؤية ورسالة المكتبات الوطنية العربية لبناء هوية رقمية لهذه المكتبات.

Keywords: *Digital Identity –Arabic National Libraries – Vision and Mission Statements - Content Analysis - Machine*

Learning Algorithms - Topic Modeling - Latent Dirichlet Allocation (LDA).

Introduction:

Currently, one of the most important keystones of digital world is the digital Identity, which has ability to distinguish entities each other's, whether these entities are individuals, organizations, locations, events, dates, or concepts in Digital and virtual world. According to digital transformation, the digital identity can be considered as initial phase to moving to digital environment for digitalizing organizations, the reasons backs to that digital identity ensures there is no confusion or distortion between similar entities, or that has the same features, so processes, electronic transactions, security, comminating, browsing, searching, retrieval, integration, exchanging which related to these entities will be accurate and secure.

According to (ENISA, 2023) A digital identity represents attributes related to an entity and is used in electronic transactions, the digital identity is unique to the context of a digital services. In this regard, an entity can be a natural person (human being), a technical user/instance (e.g., software artefact or device), or a legal entity (organization).

Digital identity is the data that internet, computer systems, cyber applications, use to represent entities which can be individuals, organizations, applications, or devices.

The Vision and Mission Statements play critical role in creating Digital Identity, where it articulates the core purpose, roles, missions, tasks, values, and strategic goals of an organization. in addition to provide a framework for decision-making and strategic planning, so, analyzing these statements can reveal the underlying themes and values that should be reflected in Arabic national Libraries digital identity.

The libraries play main roles in societies for supporting and facilitating the building of individuals' and organizations' digital identities, thereby enabling, and helping them to exist in the

digital world. Where Libraries serve as critical access points to information and digital literacy resources, which are foundational for creating and managing digital identities. They provide a wide range of digital services, including access to online databases, digital archives, and e-learning platforms, which help users develop the skills necessary to navigate and contribute to the digital environment. Additionally, libraries often offer workshops and training sessions on digital identity management, cybersecurity, and the use of various digital tools and platforms. These initiatives not only enhance individual and organizational capacities but also promote inclusivity and equal access to digital opportunities. By fostering a digitally literate community, libraries help bridge the digital divide and ensure that all members of society can participate fully in the digital age. Furthermore, libraries act as custodians of digital heritage, preserving and curating digital content that reflects the cultural and historical narratives of communities, thus contributing to a more comprehensive and inclusive digital identity landscape. Through these multifaceted roles, libraries underpin the digital transformation of society, supporting both personal and organizational growth in the digital realm (Rothfus, M, 2019).

This research proposes a novel methodology for generating and crafting digital identities for Arabic national libraries based on using machine learning algorithms to analysis contents their mission and vision statements.

Problem Statement:

A robust digital identity can enhance the accessibility of library resources, increase their visibility on global platforms, this digital existing is crucial for libraries to achieving competitive advantage, provide security, and users trust, and fulfill their mission of knowledge dissemination in the digital era.

Despite the main roles that libraries play in societies (Chang, C.H, 2007) (Chi, D.T, 2024) (Dancs, S, 2018) for supporting and facilitating building individuals and organization's digital

identity (El-Shawish, A. M. A, 2024), but many of Arabic national libraries lack to represent their self with a suitable digital identity (Abu Al-Khair, 2006).

The Main Justifications related to conduct this study backs to the vision limitations of Arabic National libraries in design their digital identities, where these identities just reflect their building or constructions by their logo, websites, and social media pages such as King Fahd national Library , Egypt national Library "Dar El-Kotob" , Qatar National Library etc... rather than reflect their main roles tasks and missions in their societies such as preserving heritage role, or digital transformation role, or knowledge access role, or cultural and educational missions in their societies, to appear on their logos, slogans, websites design, and their social media pages.

On the other hand, people can find other national libraries like Library of congress which reflect its identity by design logo as open book with waves to deliver a great message related to free access of knowledge as a main task and mission to it.

Many national libraries (include Arabic NL) are facing main challenges in creating a digital identity (Abu Al-Khair, 2006). These Challenges can be summarized as follow:

- **Strategic Alignment:** Ensuring that the digital identity of national libraries matches, aligns, and is consistent with the library's core values, vision, mission, and strategic goals is a complex task that requires careful planning and execution. To achieve this alignment between the library's digital identity and its vision and mission, AI tools and smart algorithms can be a good solution for exploring and verifying this alignment (Rothfus & Helwig, 2019).

- **Digital Identity Unawareness**: Some Arabic national libraries lack awareness of the nature of digital identity, understanding its importance, and realizing the role digital identity plays in the digital world. This lack of understanding

impedes their ability to effectively engage in marketing to their self (Allison et al., 2005)..

- **Limited Digital Identity Expertise**: The lack of experts for hinders national libraries from generating accurate digital identities that's lead to a gap between their digital presence and expected users (Rothfus & Helwig, 2019).

These challenges necessitate a structured and smart approach to creating digital identities that are not only effective but also reflective of the libraries' mission and vision, to make libraries able to be distinguishing themselves in information ecosystem.

So, before exploring the role of Arabic National libraries in enhancing digital identity for individuals and organizations in societies, main question should first be asked before:

- **How can a digital identity be created for these Arabic National libraries, to reflect their roles, tasks, missions, and future aspirations.?**

The Study Objectives:

The Main Contribution of this study is helping and guidance The Arabic National Libraries to How build a unique and a robust digital identity, depending on understanding the nature of their roles, values, and tasks in digital space.

All of that can be achieved by extracting the main identity themes from Arabic National Libraries Vision and mission Statements, based on **Analyze Mission and Vision Statements of Arabic National Libraries:** using latent Dirichlet allocation (LDA) Algorithm as content analysis algorithm to extract key themes and values from the mission and vision statements of Arabic national libraries, to understanding the libraries' strategic goals and core values.

Significance of the Study:

The significance of this study lies in its potential to contribute to creating the digital identities for Arabic national libraries based on developing a new methodology using a machine learning algorithms for content analysis to aligns digital identity

with the core values and strategic goals of these libraries. This study offers a practical solution to enhance the Arabic national Libraries digital presence. This, in turn, can lead to several important outcomes:

- **Increased Accessibility:** by digital presence based on digital identity, the digital platforms can make Arabic National libraries resources more accessible to a broader audience, including remote users, international scholars, and the public.

- **Enhanced Visibility**: A strong digital identity can raise the profile of these Arabic national libraries on global platforms, attracting more visitors, researchers, and collaborators. This increased visibility can attract international attention and foster cultural exchange and collaboration.

- **Improved User Engagement**: Interactive and user-friendly digital services can cater to the needs and preferences of modern users, enhancing their experience and satisfaction.

- **Strategic Resource Utilization**: The digital identity framework can help libraries make more strategic use of their resources by prioritizing initiatives that align with their core values and strategic goals.

Methodology:

This study explores the potential of machine learning algorithms in content analysis for extracting the main themes which reflect the philosophy, main roles, and tasks of Arabic national libraries by analyzing mission and vision statements from their websites and analyze how these attributes are functionally used to build a distinct digital identity.

Based on that, the study employed a Qualitative method to gather data from vision, mission, and objectives statements of six Arabic national libraries websites.

The practical analysis was be used for applying a machine learning LDA algorithm to content analysis for deriving main themes depending on the libraries mission statement and vision to build their digital identities.

The selection criteria will focus on Arabic National libraries that have provide access to their vision and mission statements on their official website, which was, National Library of Saudi Arabia (King Fahd National Library), National Library of UAE, Qatar National Library, National Library of Bahrain (Isa Cultural Centre), National Library of Kuwait, and National Library of Palestine. The other Arabic national libraries which have not mention or provide access their vision and mission statements by their official websites has been avoided from study population.

All the vision and mission statements have been collected form official Arabic National Libraries websites in 2024.

Main steps to conduct this analysis are:

▪ **Data Collection**: Gathering mission and vision statements from websites study population (Arabic national libraries official websites).

▪ **Preprocessing**: Cleaning and preparing data for analysis, including tokenization, stop-word removal, and stemming.

▪ **Content Analysis**: Applying algorithms such as Latent Dirichlet Allocation (LDA) to identify core themes in the statements.

▪ **Digital Identity Framework: Developing a framework for translating the extracted themes and values into digital identity ideas for these national libraries, where these themes can be used for creating logos, slogans, such website design, social media presence, and digital content strategies to these libraries.**

Literature Review:

- Understanding Digital Identity:

Many Definitions has been presented to digital identity in various domains (Windley, P. J., 2005) (Allison, A, 2005) (Sullivan, C, 2011) (Laurent, M, 2015) (Atick, J, 2016).

National Institute of Standards and Technology (NIST) confirmed it is difficult to land on a single definition of digital identity without context or domains (Grassi, P. etc. 2017).

According to (Sedlmeir, J, 2021) to create a common basis for discussion about digital identities in various domains, it is important to first define many terms are related to the digital identity.

For this context (ISO, 2019) defined the as "**Entity**" can be defined as "an item relevant for the purpose of operation of a domain, it has recognizably distinct existence", and the entity can have a physical or a logical embodiment, for example, it can be a person, an organization. While the "**Identity**" has been defined as a set of attributes, that related to an entity, where the entity can have more than one identity, and the same time a several entities can have the same identity (like the name of category which belonging to it many entitles).The "**Attribute**" is a characteristic or property of an entity, for example: an entity type, address information, telephone number. While the "**Domain**" is an attribute value was created or its value has been (re)assigned, the domain of origin can be provided as meta data for an attribute. The "**Identifier**" is an attribute or set of attributes that uniquely characterizes an identity in a domain. The "**Identification**" is a process of recognizing an entity in a particular domain as distinct from other entities.

Depending on these terms, **The Digital Identity** is "a set of attributes related to an entity (such as a person, organization, or device) that is used within a given context. These attributes uniquely describe the entity within that context, enabling identity management systems to distinguish and manage the entity effectively" (European Commission, 2020) (Sedlmeir, J, 2021).

So, the digital Identities **can be defined according to libraries perspective**, "the online persona and digital presence of an institution or individual". For national libraries, this

encompasses their website, social media profiles, digital content, and online interactions with users, where a strong digital identity reflects the core values, mission, and strategic goals of the library, ensuring consistency and coherence across all digital platforms.

History of the Digital Identity backs to 1999, where the European Union published Directive 1999/93/EC, which provided the first framework for electronic signatures. Which can be considered as a first phase to creating digital identity for individual (European Union, 1999).

The Importance of Digital Identity back to establish trust and understanding among stakeholders in an ecosystem. If stakeholders do not trust the identity of their peers, the data held in the solution will be deemed unreliable, the overall ecosystem will lose its effectiveness (Škurtan, M, 2020).

According to (Sullivan, C, 2019) the UN mandating nations to provide legal and digital identity for all organizations and individuals by 2030, this is considering a goal underpins Sustainable Development Goals be achieved by the UN member nations.

Main Properties of Digital Identity According to (Rundle, M, 2007) are: Essentially social – Subjective -Valuable – Referential – Composite – Dynamic - Potentially.

Key **Components of Digital Identity** has been illustrated by National Institute of Standards and Technology (NIST) which are three components two of them are basic and a third optional component:

- **Component One (essential): Identity proofing and enrolment:** Identity proofing answers the question, "Who are you?" and refers to the process by which an identity service provider (IDSP) collects, validates, and verifies information about the entities and resolves it to a unique individual within a given context.

- **Second Component (essential): Authentication and identity lifecycle management**: An authenticator is something the claimant possess and controls—typically, a cryptographic module, one time code generator or password—that is used to authenticate (confirm) the claimant.

- **Third Component Three (optional): Portability and interoperability mechanisms:** Digital ID systems can—but need not--include a component that allows proof of official identity to be portable. (FATF, 2020).

Despite importance of Digital Identities, it is facing **Significant Challenges**: **Security and Privacy Concerns:** One of the most significant challenges in digital identity management is ensuring security and privacy. **Interoperability Issues:** refers to the ability of different systems and organizations to work together. **Entity Adoption and Trust:** Gaining user trust and encouraging widespread adoption of digital identity solutions is challenging. **Regulatory and Compliance Issues:** Different countries have varying regulations regarding digital identity management. **Technological Limitations:** current technologies still face limitations, such as the lack of a universally accepted standard for digital identity. **Inclusivity and Accessibility:** Lack of access to technology, digital skills, and infrastructure can exclude significant portions of the population from digital identity benefits. **Ethical and Social Implications** Issues like surveillance, discrimination, and loss of anonymity are concerns that need to be addressed (Cavoukian, A, 2013) (Windley, P. J, 2005) (Cameron, K, 2005) (European Union, 2016) (Zyskind, G., & Nathan, O, 2015) (Gelb, A., & Clark, J, 2013) (Lyon, D, 2009).

The Digital Identity has many **Tools** can be shown through it, according to (Chaffey, D, 2020) (Score, 2024) the best and robust tools to show the digital identities for organizations are: **Website Design**: A well-designed website that is user-friendly, accessible, and reflective of the library's mission and values.

Social Media Presence: Active engagement on social media platforms to reach and interact with audience. **Digital Content**: High-quality digital collections, virtual exhibits, online catalogs, and other digital resources. (Chaffey, D, 2020) (Score, 2024).

Many **Digital Identity Standards** have been issued (ENISA, 2023) (World Bank, 2018) (FATF, 2020) (New-Zealand government, 2018) (Grassi, P. A, 2017), IDunion. (2024) where it can cover several domains, the main aims for these standards were variant such as describing Digital identity policies (ENISA, 2023).

- **Content Analysis Algorithms:**

Content analysis has been **Defined** as a systematic, replicable technique and method for analyze text data by compressing many words of text into fewer content categories based on explicit rules of coding to identify patterns, themes, and meanings, as well defiened as, the study of the content with reference to the meanings, contexts and intentions contained in messages (Prasad, B, 2008) (Krippendorff, K,2018) (Berelson, 1952) (Stemler, S, 2000) (Weber, 1990).

According to (Stemler, S, 2000), Content Analysis **History** backed to 75 years old. development of content analysis as a full-fledged scientific method took place during World War II when the U.S. government sponsored a project under the directorship of Harold Lasswell to evaluate enemy propaganda. In 1952 the term has been coined by Bernard Berelson when published his primer book Content analysis in Communication Research (Berelson, 1952), and the content analysis term had been inserted into Webster's Dictionary of English language since 1961.

According to (Prasad, B. D, 2008) The Main **Objective** of content analysis is to convert recorded "raw" phenomena into data, which can be treated in essentially a scientific manner so that a body of knowledge may be built up.

Regarding to Content Analysis **Uses** & **Applications**, Holsti (1969) definition explained that content analysis techniques isn't restricted to textual analysis, but may be applied to other areas such as coding (Wheelock, Haney, & Bebell, 2000), multimedia (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999) (GAO, 1996), about the scientific domains, Prasad (2008) illustrated that the content analysis It is not limited only to social science research, that was emphasized by (Damyanov, M, 2023) content analysis can be expanded to be involved into applications in anthropology, communications, history, linguistics, literary studies, marketing, political science, psychology, and sociology, AI studies and other disciplines.

There are several **Types** of content analysis (Forman, J, 2007) (Harwood, T, 2003) (Drisko, J, 2016) all sharing the central feature of systematically categorizing textual data to make sense of it (Miles & Huberman, 1994). However, they differ, in the ways they generate categories and apply them to the data, and how they analyze the resulting data. According to (CUMSPH, 2024) There are two general types of content analysis: conceptual analysis and relational analysis. **Conceptual analysis** determines the existence and frequency of concepts in a text. **Relational analysis** develops the conceptual analysis further by examining the relationships among concepts in a text. Each type of analysis may lead to different results, conclusions, interpretations, and meanings.

According to (Delve, Ho, 2023) the **Thematic Analysis** is the most popular Conceptual analysis subtype. (Joffe, 2012) points out that thematic analysis originally emerged from content analysis, but it developed into a separate approach with its own unique research goals. **Thematic Analysis** is a method for identifying and analyzing patterns of meaning in a dataset (Braun & Clarke, 2006).

Conducting content analysis on text data, the experts in this field (Bos & Tarnai, 1999) (Neuendorf, 2017) (Riffe, Lacy &

Fico, 2014) (Weber, 1994) (White & Marsh, 2006) (Zamith & Lewis, 2015) broadly agreed on many commons' steps among them to execute content analysis where (CUMSPH, 2024) formulated them in six steps, which are: 1) Determine the types of relationships and level of analysis (word, phrase, sentence, themes). (2) Categorize the text and code for words or patterns. (3) Explore relationships between concepts, considering strength, sign, and direction. (4) Code the relationships, differentiating from conceptual analysis. (5) Perform statistical analyses to explore differences or relationships among variables. (6) Map representations like decision maps and mental models. According to (Paperpile, 2024) conducting Thematic analysis is include six steps:1) Familiarization. 2) Generating codes. 3) Generating themes. 4) Reviewing themes. 5) Defining and naming themes, 6) Creating the report.

Technologies of Content and Thematic analysis (Krippendorff, K, 2018) have evolved significantly with advancements in computational methods, artificial intelligence, and machine learning. The most popular technologies are using for that are:

Natural Language Processing (NLP) (Pirnau, M, 2024) (Crowston, K,2012): NLP techniques are used to process and analyze large volumes of text data. They enable the extraction of meaningful information from text, such as themes, sentiments, and patterns.

Machine Learning Algorithms: (Scharkow, M, 2013) (Fussell, 2022) (Nastase, 2007) are employed to identify patterns and make predictions based on the analyzed content. Classification algorithms e.g., support vector machines, decision trees and clustering algorithms e.g., k-means, hierarchical clustering are widely used.

Text mining (Aureli, S, 2017**) (**Vijayarani, S, 2015**) (**Patel, F, 2012**):** involves extracting useful information from text. It includes processes such as information retrieval, lexical analysis, pattern recognition, and data mining.

Topic modeling is a statistical method for discovering abstract topics within a collection of documents. It helps in identifying the main themes or topics discussed in the content. Topic models were originally developed as a means of automatically indexing, searching, clustering, and structuring large corpora of unstructured and unlabeled documents (Jelodar, H, 2019) (Chen T-H, 2016).

Latent Dirichlet Allocation (LDA): LDA is a topic modeling technique that identifies topics within a set of documents.. Latent Dirichlet Allocation (LDA) is a generative probabilistic model that assumes documents are mixtures of topics and topics are mixtures of words. It aims to identify the hidden topics that pervade a collection of documents and to determine the distribution of topics within each document and the distribution of words within each topic. Latent Dirichlet Allocation (LDA) is commonly used for topic modeling It can be used to extract key themes from mission and vision statements (Jelodar, H, 2019) (Hoffman, M, 2010).

Gensim: A Python library for topic modeling and document similarity analysis. It is useful for unsupervised learning tasks such as identifying the topics in a corpus. (Řehůřek, R, 2024)

Sentiment Analysis (Tan, K, 2023**) (**Ligthart, 2021**) (**Taherdoost, H, 2023**)**: Sentiment analysis involves determining the emotional tone of text. It can help in understanding the sentiment expressed in mission and vision statements, whether positive, negative, or neutral.

National libraries have traditionally served as custodians of a nation's cultural heritage. They preserve and maintain extensive collections of manuscripts, books, periodicals, maps, and other documents of historical and cultural significance. In the Arab world, this role is particularly crucial due to the region's rich historical and cultural legacy. Beyond preservation, national libraries play a significant role in promoting knowledge and education. They serve as centers for learning, offering resources

and services that support educational institutions, researchers, and the public (Rubin, R. E, 2020).

Content analysis has been widely used in library and information science to study various aspects of libraries (Armann-Keown, 2020), including user behavior, service quality, and organizational communication. In the context of digital identity, content analysis can help libraries understand how their mission and values are perceived and communicated through their digital platforms. It provides a systematic approach to aligning digital content and strategies with the core values and strategic goals of the library (Abdullah, Z, 2022).

Several national libraries worldwide have successfully developed strong digital identities by leveraging content analysis and strategic planning.

• **The British Library**: The British Library has an extensive digital presence, including digitized collections, virtual exhibitions, and active social media engagement. Their digital strategy is closely aligned with their mission of making their collections accessible to everyone (British Library, 2020).

• **Library of Congress**: The Library of Congress has implemented numerous digital initiatives, including the digitization of historical documents and the development of interactive digital platforms. Their digital identity reflects their commitment to preservation and public access (Library of Congress, 2023).

• **National Library of Australia**: The National Library of Australia has developed a strong digital identity through initiatives such as Trove, an online portal that provides access to a vast array of digital resources. Their digital strategy focuses on user engagement and accessibility (National Library of Australia, 2023).

In many Arab countries, national libraries act as hubs for literacy programs, educational workshops, and public lectures,

fostering a culture of learning and intellectual growth (Wahdan, D. M, 2020).

The digital presence of a library is no longer a supplementary feature but a core component of its identity and service delivery. A well-crafted digital identity can significantly enhance a library's ability to reach wider audiences, provide innovative services, and remain relevant in a rapidly changing digital landscape (Meriam, M, 2023).

Developing The Model:

As mentioned previously, Content analysis involves systematically analyzing text data to identify patterns, themes, and insights. For Arabic national libraries, this process will be used to extract key themes and values from their mission and vision statements, which will then inform the development of their digital identity. The content analysis process includes several key steps: data collecting, preprocessing, theme extraction, sentiment analysis, and mapping themes.

Mission and vision statements articulate the core purpose, values, and strategic goals of an organization. For Arabic national libraries, these statements provide a foundation for their digital identity.

Many techniques have been used to creating the digital identity, such as Blockchain, FIDO Universal Authentication Framework, etc. but the machine learning (ML) algorithms, especially content analysis techniques are the most powerful tools is using for creating a digital identity (World Bank, 2018).

Machine Learning Content analysis algorithms, especially topic modeling algorithms like latent Dirichlet allocation LDA algorithm will be used to systematically analyze Arabic national libraries mission and vision statements for extracting key themes and values that should be reflected in the library's digital identity.

The extracted themes and values will be translated into actionable digital strategies, this ensures a robust presence for

these libraries on the digital world by designing a professional website that reflects the library's mission, creating engaging digital content, and developing a robust social media presence. The roadmap to developing Digital Identity Framework for Arab national libraries has many steps, as follows:

1. **Data Collection:**
- **Data Sources & Data Acquisition:**
o **The primary data for this study will be the mission and vision statements of Arabic national libraries. These statements are typically found on the libraries' official websites. So, the Arabic national Libraries official websites are the only source that have been depended on it for capturing mission and vision statements for these libraries**.
o Manual data collection methods have been used to gather the mission and vision statements from the Arabic National libraries' websites for Ensuring the accuracy and completeness of the collected data.

2. **Preprocessing:**
Preprocessing is a crucial step in content analysis, as it prepares the text data for further analysis. It involves cleaning the text, tokenizing it, and transforming it into a format suitable for extracting meaningful insights.
- **Text Cleaning:** The raw text data collected from the mission and vision statements will undergo preprocessing to remove any noise and prepare it for analysis to make standardizing text data. This includes:
▪ **Normalization**: Converting all text to lowercase to ensure consistency. so that "Library" and "library" are treated as the same word.
▪ **Punctuation Removal:** Remove punctuation marks that do not contribute to the semantic meaning of the text, such as commas, periods, and exclamation marks.
▪ **Special Character Removal:** Removing special characters and numbers that are irrelevant to the text analysis.

▪ **Handling Arabic Text:** Addressing specific challenges related to Arabic text, such as dealing with diacritics and different word forms.

- **Tokenization:** involves splitting the text into individual words or phrases (tokens). This process helps in analyzing the text at a granular level. Different tokenization techniques, such as word-level and phrase-level tokenization, has been employed based on the specific requirements of the Mission and Vision analysis.

▪ **Word-level Tokenization:** Splitting text into individual words, which is useful for basic analysis, and frequency counts.

▪ **Phrase-level Tokenization:** Break text into meaningful phrases or n-grams (e.g., bi-grams, tri-grams) to capture more context, for example, "digital library" and "cultural heritage" can be treated as single tokens.

- **Stop-word Removal**: Eliminating common words that do not contribute meaningful information.

▪ **Common Stop-words**: stop-words (common words e.g., "في", "من", "و") have been removed from the text as they do not carry significant meaning and can clutter the analysis.

▪ **Customized Stop-word List:** Creating a stop-word list tailored to the context of the mission and vision statements and specific to the context of national libraries and cultural heritage., ensuring that only non-informative words are removed.

- **Stemming and Lemmatization:** Stemming and lemmatization standardize different forms of the same word to ensure consistency in meaning by reducing words to their base or root form to standardize variations, Both Stemming & Lemmatization techniques will be applied to standardize the mission and vision text and ensure that different forms of the same word are treated as a single entity.

▪ **Stemming:** involves reducing words to their base or root form (e.g., "الحفاظ" to "حفظ").

▪ **Lemmatization**: Applying to convert words to their root forms, which is particularly important for handling the inflectional morphology of the Arabic language. (e.g., "better" to "good", "الشبكة العنكبوتية" to "الويب").

- **Data Annotation**: Data annotation is a critical step in preparing text data for content analysis. It involves labeling the data with meaningful tags that reflect the content, context, and sentiment. for this study, the researcher annotates important data manually for training algorithm and validation purposes (Table1), for Example: according to thematic annotation:

o **Original Text**: "*The National Library of Egypt aims to provide access to cultural heritage and support educational initiatives."*

o **Annotated Text**: *"The National Library of Egypt "[**Access**] The library aims to provide access to cultural heritage [**Cultural Heritage**] and support educational initiatives [**Education**]".*

Annotation Category	Example
Preservation	"Our mission is to preserve…"
Education	"The library offers free online courses"
Digital	"To transform into digital …"
Accessibility:	"The library provides access to rares…"
Innovation	"To be a beacon of innovation,..."
Collaboration	"To be a hub of collaboration…'
Trust	"To be trusted guardian of knowledge"
Leadership	"To be a global leader in library …"

Table1: Contains List of Vision & Mission Annotation.

3. Content Analysis Model (Theme Extraction):

Suggested model has been created to identify core themes and sentiments in the vision and mission statements, that was had been done depending on Natural Language Processing (NLP) techniques which are used to process and analyze large volumes of text data. They techniques enable the extraction of

meaningful information from text, such as themes, sentiments, and patterns. In this Model, theme extraction involves identifying and categorizing the main topics and ideas expressed in the text data vision and mission statements of the Arabic National Libraries.

- **The Latent Dirichlet Allocation (LDA) Algorithm:** Applying LDA to creating and applying topic modeling for extract key themes from mission and vision statements. Topic modeling is a methodology for discover abstract themes within the text data.

▪ **Topic Modeling:** Use LDA to identify clusters of words that frequently occur together, revealing underlying themes in the mission and vision statements. For example, Applying LDA might reveal topics such as:

o Library Vision 1: ["**preserve**", "**heritage**", "**cultural**", "**tradition**"],

o Library Vision 2: ["**accessible**", "**knowledge**", "**education**", "**citizens**"].

▪ **Parameter Tuning**: Determine the optimal number of topics by experimenting with different parameters and using coherence scores to evaluate the quality of the topics contained in vision and mission statements.

▪ **Interpretation of Topics**: Analyze the topics to identify the core themes and values expressed in the vision and mission statements. by thisi step the model labeled each topic based on the most representative words.

- **Frequency Analysis:**

▪ **Term Frequency (TF):** Calculate the frequency of individual words and phrases to identify the most commonly occurring terms in the mission and vision statements.

▪ **Term Frequency-Inverse Document Frequency (TF-IDF):** Use TF-IDF to identify terms that are important within a specific document but less common across the entire corpus, highlighting unique themes.

For Example, Frequency analysis shows that **"Heritage"** and **"Accessible"** are among the most frequent terms. So, highlighting preservation and accessibility as key themes.

- **Word Clouds and Visualizations:**
▪ **Word Clouds:** Creating word clouds to visualize the most frequent terms and their prominence in the mission and vision statements.

▪ **Graphs and Charts:** Using bar charts, histograms, and other visual tools to display the frequency of key terms and themes in vision and mission, aiding in the interpretation of the data.

4. **Sentiment Analysis:**
Sentiment analysis assesses the emotional tone of the text to understand the libraries' emphasis on various aspects such as innovation, inclusivity, and accessibility.

- **Sentiment Classification:**
▪ Sentiment Model: since training data is limited, the study utilizing pre-trained models such as Watson Discovery Model to classify the sentiment of the vision and mission text to explore a historical aspect, future aspects.

5. **Digital Identity Framework (Mapping Theme)**
By Suggested Model, the Mapping themes framework involves categorizing the identified themes into broader categories, prioritizing them, and generating actionable insights for creating Digital identity Framework for each Arabic National Library, to help them to exploit these themes in developing and design their own digital identities tools (like Official websites – Social network pages – strategic plans) for reflecting their natures, this framework is depending on four steps which are:

- **Categorization of Themes:**
▪ **Broad Categories:** Grouping identified themes into broader categories such as accessibility, preservation, education, community engagement, and innovation.

▪ **Sub-themes:** Further dividing these broad categories into sub-themes for more detailed analysis. For example,

"education" could include sub-themes like "research support," "public programs," and "literacy initiatives."

- **Prioritization of Themes:**

▪ **Frequency and Importance:** Prioritizing themes based on their frequency and importance in the mission and vision statements. Themes that appear more frequently or are emphasized strongly should be given higher priority.

▪ **Strategic Relevance:** Considering the strategic relevance of each theme in the context of the library's goals and objectives. Themes that align closely with the library's strategic priorities should be prioritized.

- **Theme Correlation and Relationships:**

▪ **Co-occurrence Analysis:** Analyzing the co-occurrence of themes to identify relationships and patterns. For example, how often do themes like "innovation" and "accessibility" appear together?

▪ **Thematic Networks:** Creating thematic networks to visualize the relationships between different themes and how they interconnect within the mission and vision statements.

For example: Themes are categorized into "Preservation" and "Accessibility." These themes are prioritized based on their frequency and strategic relevance. While Co-occurrence analysis shows a strong relationship between "preserve" and "heritage," indicating a focused effort on cultural preservation.

- **Actionable Insights:**

▪ **Strategic Insights:** Using the identified themes to generate strategic insights that can inform the development of digital identity initiatives. For example, if "community engagement" is a prominent theme, consider creating interactive community-focused digital content.

▪ **Operational Plans:** Translating the thematic insights into operational plans and actions that align with the library's mission and vision. This could include developing specific

digital programs, content strategies, and user engagement initiatives.

6. **Tools and Techniques**

- **Software and Tools:**

▪ **NLP Libraries:** Using NLP libraries such as NLTK, spaCy, and Gensim for text preprocessing, tokenization, and topic modeling.

▪ **Data Visualization Tools:** Employing data visualization tools such as Matplotlib, Seaborn, and WordCloud for creating visual representations of the analysis.

▪ **Machine Learning Frameworks:** Utilizing machine learning frameworks like Scikit-learn and TensorFlow for sentiment analysis and model training.

- **Collaboration and Documentation:**

▪ **Collaborative Platforms:** Using collaborative platforms like Jupyter Notebooks or Google Colab for sharing code, results, and visualizations with the team.

▪ **Documentation:** Maintaining thorough documentation of the analysis process, including methodologies, parameter settings, and interpretation of results to ensure transparency and reproducibility.

Model Implementation & Case Studies:

The Suggested Model aims to extract key themes and values from the mission and vision statements of Arab national libraries, according to these phases:

- Data Collection.
- Preprocessing.
- Annotation.
- Theme Extraction.
- Practical Applications (Mapping Themes to Digital Elements):

Case 1: UAE National Library[2]:

Data Collection:

▪ **Vision Statemen:** "Connecting, inspiring, enriching knowledge societies."

▪ **Mission Statement:** "Be the custodian of our nation's documentary heritage through the collection, preservation, and access to enable knowledge societies."

▪ **Strategic Objectives:** "Become an esteemed reference entity for collections of national significance through building, safeguarding, and disseminating intellectual material.", "Establish a sustainable global knowledge beacon." Etc.

Preprocessing:

▪ **Original Vision Statement:** "Connecting, inspiring, enriching knowledge societies."

▪ **Normalized Vision Statement:** "connecting inspiring enriching knowledge societies."

▪ **Original Mission Statement:** "Be the custodian of our nation's documentary heritage through the collection, preservation, and access to enable knowledge societies."

▪ **Normalized Mission Statement:** "be the custodian of our nation's documentary heritage through the collection preservation and access to enable knowledge societies." (Figure 1).

Annotation:

▪ **Original Vision Statement:** "Connecting, inspiring, enriching knowledge societies."

▪ **Annotated Text:** "connecting [**Collaboration**] inspiring [**Innovation**] enriching knowledge societies [**Education**]."

▪ **Original Mission Statement:** "Be the custodian of our nation's documentary heritage through the collection, preservation, and access to enable knowledge societies."

▪ **Annotated Text:** "connecting [**Collaboration**] inspiring [**Innovation**] enriching knowledge societies [**Education**]."

Figure 1: The Output after applying Normalization on the UAE National Library Vision by the Model.

Theme Extraction

Using developed topic model (Latent Dirichlet Allocation Algorithm) for extracting main themes form vision and mission statements, then applying frequency analysis to identify key themes across all vision, mission, and values statement. (Figure 4).

▪ **Key Themes:** Preservation, Accessibility, Education, Innovation, Collaboration, Trust, Openness, Learning, Pro-activeness (Figure 2).

Practical Applications:

The extracted themes and values from the content analysis will be translated into actionable digital strategies to exploit these Themes into Digital Identity Application, to build strong digital identity and robust online presene for UAE National Library. These include, design website to reflect **Trust** Theme, and **Accessibility** to all users, **Innovation**, **Collaboration**, and other themes. This mapping includes the using of social media to promote heritage **Preservation** projects and initiatives. and

include also creating digital content reflects **Education** themes by produce online courses, webinars, and educational materials.

Figure 2: Output from Applying LDA Model on The UAE's Vision and Mission to Extract Main Themes & their frequency,

Case 2: King Fahd National Library[3]:

Data Collection:

▪ **Vision Statement:** "Leadership in preserving the national intellectual product and the Arab and Islamic heritage and excellence in organizing and making it available to build a knowledge society."

▪ **Mission Statement:** "Preserving information and making it available by employing advanced technologies, leading library and information work, and organizing national knowledge initiatives through partnership with cultural institutions to serve all segments of society and advance it cognitively."

Preprocessing:

▪ **Normalized Vision:** "leadership in preserving the national intellectual product and the arab and islamic heritage and excellence in organizing and making it available to build a knowledge society."

▪ **Normalized Mission:** "preserving information and making it available by employing advanced technologies, leading library and information work, and organizing national knowledge initiatives through partnership with cultural institutions to serve all segments of society and advance it cognitively."

Annotation:

• **Annotated Vision:** "leadership [**Leadership**] in preserving [**Preservation**] the national intellectual product and the arab and islamic heritage [**Heritage**] and excellence [**Excellence**] in organizing and making it available [**Accessibility**] to build a knowledge society."

• **Annotated Mission:** "preserving [**Preservation**] information and making it available [**Accessibility**] by employing advanced technologies [**Digital**], leading [**Leadership**] library and information work, and organizing national knowledge initiatives through partnership [**Collaboration**] with cultural institutions to serve [**Pro-activeness**] all segments of society and advance it cognitively."

Theme Extraction:

• **Key Themes**: Preservation, Accessibility, Leadership, Digital, Pro-activeness, Collaboration, Excellence, Heritage (Figure 3).

Practical Application:

• **Website Design**: Emphasize sections on cultural heritage preservation and accessible knowledge resources.

• **Social Media Strategy**: Use social media to promote accessibility initiatives and heritage preservation projects.

• **Digital Content Creation**: Develop virtual exhibits and digital archives focused on Saudi heritage.

Figure 3: Extracted Themes and their Frequencies as Output from LAD Model to KSA National Library

Case 3: Qatar National Library[4]:

Data Collection:

▪ **Vision Statement:** "To be one of the world's preeminent centers of learning, research and culture; a guardian of the region's heritage; and an institution that promotes imagination, discovery and the nourishment of the human spirit."

▪ **Mission Statement:** "Qatar National Library aims to bridge knowledge gaps, support lifelong learning, and inspire cultural engagement through innovative digital services."

Preprocessing:

▪ **Normalized Vision**: "to be one of the world's preeminent centers of learning, research and culture; a guardian of the region's heritage; and an institution that promotes imagination, discovery and the nourishment of the human spirit."

▪ **Normalized Mission:** "qatar national library aims to bridge knowledge gaps, support lifelong learning, and inspire cultural engagement through innovative digital services."

- **Annotation:**
- **Annotated Vision:** "to be one of the world's preeminent [**Leadership**] centers of learning [**Learning**], research [**Learning**] and culture [**Culture**]; a guardian of the region's heritage [**Heritage**]; and an institution that promotes imagination, discovery [**Inspiration**] and the nourishment of the human spirit."
- **Annotated Mission:** "qatar national library aims to bridge knowledge gaps [**Education**], support lifelong learning [**Education**], and inspire cultural engagement [**Cultural Engagement**] through innovative digital services [**Digital Innovation**]."
- **Theme Extraction:**
- **Key Themes**: Education, Cultural Engagement, Digital Innovation, Inspiration, Leadership, and Heritage. (Figure 4).
- **Practical Application:**
- **Website Design**: Highlight digital services, educational programs, and cultural events.
- **Social Media Strategy**: Engage users with interactive digital content and cultural programs.
- **Digital Content Creation**: Create online courses, webinars, and digital storytelling projects.

Figure 4: Extracted Themes and their Frequencies as Output from LAD Model to QNL.

Case 4: Kuwait National Library [5]:

Data Collection:

▪ **Vision Statement:** "To consolidate the role of the library in preserving the history of Kuwait, illuminating its present, and safeguarding national memory."

▪ **Mission Statement:** "To be a beacon of culture and knowledge, and a national civil center that encompasses and safeguards scientific, creative, and literary production."

Preprocessing:

▪ **Normalized Vision**: "to consolidate the role of the library in preserving the history of kuwait, illuminating its present, and safeguarding national memory."

▪ **Normalized Mission:** "to be a beacon of culture and knowledge, and a national civil center that encompasses and safeguards scientific, creative, and literary production."

• **Annotation:**

▪ **Annotated Vision:** "to consolidate [**Leadership**] the role of the library in preserving [**Preservation**] the history [**Heritage**] of kuwait, illuminating its present, and safeguarding national memory[**Heritage**]."

• **Annotated Mission:** "to be a beacon [**Leadership**] of culture [**Culture**] and knowledge [**Knowledge**], and a national civil center that encompasses and safeguards scientific, creative [**Creativity**], and literary production [**Literary Production**]."

Theme Extraction:

• Key Themes: Leadership, Cultural, Digital, Preservation , Leadership, Literary Production, and Heritage. (Figure 5).

Practical Application:

• **Website Design:** Preservation of History: Create dedicated sections for historical archives and initiatives aimed at preserving the history of Kuwait.

• **Social Media Strategy:** Culture and Knowledge: Share updates on cultural events, knowledge dissemination programs, and educational resources.

- **Digital Content Creation:** Scientific and Literary Production: Create interactive digital content showcasing scientific research, creative works, and literary production.

Figure 5: Extracted Themes and their Frequencies as Output from LAD Model to Kuwait National Library.

Case 5: Bahrain National Library [6]:

Data Collection:

- **Vision Statement:** "We would like to see Isa Cultural Center a replica of Bahrain in the dawn of its civilization as a Center for knowledge to the civilized world, absorbing all new and useful human inputs to remain a symbol of Bahrain's progress and prosperity over the time".

- **Mission Statement:** "Comes to spread Bahraini culture at the level Local and international, providing distinguished and effective cultural services to segments of society of various groups and levels, and being open to global cultures."

Preprocessing:

- **Normalized Vision**: "we would like to see isa cultural center a replica of bahrain in the dawn of its civilization as a center for knowledge to the civilized world, absorbing all new and useful human inputs to remain a symbol of bahrain's progress and prosperity over the time".

- ▪ **Normalized Mission:** "comes to spread bahraini culture at the level local and international, providing distinguished and effective cultural services to segments of society of various groups and levels, and being open to global cultures."
- • **Annotation:**
- ▪ **Annotated Vision:** "we would like to see isa cultural center a replica of bahrain in the dawn of its civilization [**Civilization**] as a center for knowledge [**Knowledge**] to the civilized world, absorbing all new and useful human inputs to remain a symbol of bahrain's progress [**Progress**] and prosperity [**Prosperity**] over the time"
- ▪ **Annotated Mission:** "the mission of the isa cultural center comes to spread [**Leadership**] bahraini culture [**Culture**] at the level local and international [**Global**], providing [**Accessibility**] distinguished and effective cultural services to segments of society of various groups and levels, and being open to global cultures [**Culture**]."

Theme Extraction:
- • **Key Themes**: Cultural, Global, Accessibility, Knowledge, Progress, Prosperity, Leadership, and Civilization. (Figure 6).

Practical Application:
- • **Website Design:** Cultural Heritage and Civilization: Create dedicated sections for showcasing Bahrain's cultural heritage and historical archives.
- • **Social Media Strategy:** Highlight projects and initiatives that contribute to Bahrain's progress and prosperity.
- • **Digital Content Creation:** Develop virtual exhibits and digital archives focused on Bahrain's cultural heritage and civilization.

```
import gensim
from gensim import corpora

# Bahrain National Library
texts = [
    " see isa cultural center replica bahrain  dawn  civilization  center  knowledge  civilized world absorbing  useful
    "Comes  spread Bahraini culture   level Local  international, providing distinguished  effective cultural services  segmen
]

# Tokenize and prepare dictionary and corpus
tokenized_texts = [text.split() for text in texts]
dictionary = corpora.Dictionary(tokenized_texts)
corpus = [dictionary.doc2bow(text) for text in tokenized_texts]

# LDA model
lda_model = gensim.models.LdaModel(corpus, num_topics=2, id2word=dictionary, passes=15)

# Print the topics
topics = lda_model.print_topics(num_words=9)
for topic in topics:
    print(topic)
```

```
(0, '0.039*"cultural" + 0.039*"center" + 0.024*"society" + 0.024*"various" + 0.024*"being" + 0.024*"global" + 0.024*"open" + 0.
(1, '0.024*"center" + 0.024*"cultural" + 0.024*"useful" + 0.024*"bahrain's" + 0.024*"bahrain" + 0.024*"progress" + 0.024*"symbo
```

Figure 6: Extracted Themes and their Frequencies as Output from LAD Model to Bahrain National Library.

Case 5: Palestine National Library [7]:

Data Collection:

▪ **Vision Statement:** "Reaching the Palestinian cultural intellectual production to a distinguished stature, to be recognized globally; officially and popularly".

▪ **Mission Statement:** "The Palestinian intellectual and cultural production and heritage are a genuine living part of the human production that contributes to the development of global civilization which deserves special care and effective international protection."

Preprocessing:

▪ **Normalized Vision**: "reaching the palestinian cultural intellectual production to a distinguished stature, to be recognized globally; officially and popularly".

▪ **Normalized Mission:** "the palestinian intellectual and cultural production and heritage are a genuine living part of the human production that contributes to the development of global civilization which deserves special care and effective international protection."

- **Annotation:**
▪ **Annotated Vision:** "reaching [**Accessibility**] the palestinian cultural [**Culture**] intellectual production [**Knowledge**] to a distinguished [**Identity**] stature, to be recognized globally [**Global Recognition**]; officially and popularly".
▪ **Annotated Mission:** "the palestinian intellectual [**Knowledge**] and cultural production and heritage [**Heritage**] [**Preservation**] are a genuine living part of the human production that contributes to the development of global [**Global Recognition**] civilization which deserves special care and effective international protection."

Theme Extraction:

- **Key Themes**: Cultural, Global Recognition, Accessibility, Knowledge, Preservation, Heritage, Identity and Civilization. (Figure 7).
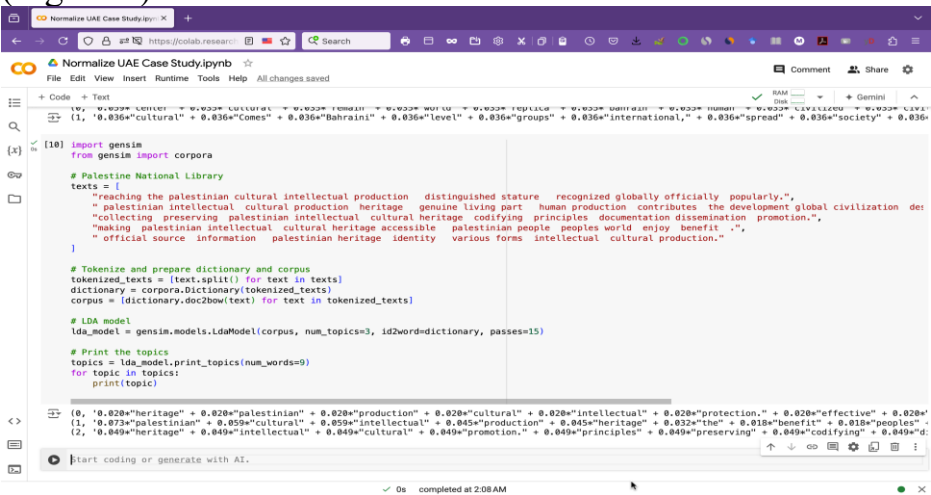

Figure 7: Extracted Themes and their Frequencies as Output from LAD Model to Palestine National Library.

Practical Application:

- **Website Design:** Heritage and Cultural Production: Create dedicated sections for Palestinian cultural and intellectual heritage and production, including digital archives and collections.

- **Social Media Strategy:** Global Recognition: Share updates on international recognition and collaboration projects.
- **Digital Content Creation:** Create content that emphasizes the global significance and recognition of Palestinian cultural production.

Discussion & Conclusion:

In conclusion, the study presented comprehensive overview of the Digital Identity and the critical roles that plays, and the opportunities and challenges associated with developing a strong digital identity for Arabic national libraries. where developing a strong digital identity has significant implications for Arabic national libraries. It affects their ability to preserve cultural heritage, increase accessibility, engage with the community, and align with international standards.

Study aimed to helping the Arabic National Libraries to build their digital identities not by show their building in logos, websites, and social network pages, but by deliver their message and main roles in society, such as preserving heritage, enhancing digitalization, knowledge supply and support cultural & educational aspects.

Study aims too, verify the potential of content analysis machine learning algorithms in creating and enhancing digital identities for Arabic National Libraries, depending on their mission statement and objectives.

Study explored that applying content analysis to the mission and vision statements of Arabic national libraries is an objective way to extract main themes to these libraries which are central to the libraries' identities.

To achieve that, these libraries should be know what's the main themes should be reflect in their digital identity design.

this study developed an objective method to determine main themes of these libraries, depending on Machine Learning algorithms *Latent Dirichlet Allocation LDA*, by analysis these libraries vision and mission statements to extract main themes,

to help libraries focus on it when designing their websites, logos, and social media posts.

Before creating and applying the study model, the study provided a comprehensive and detailed roadmap for analyzing the mission and vision statements of Arabic national libraries. that's includes steps for preprocessing text, extracting themes, performing sentiment analysis, and mapping themes to generate actionable insights.

One of the most importance study findings was the topic modeling with LDA algorithm is a powerful methodology for discovering hidden themes in Arabic National Libraries vision and mission statements. By identifying patterns and topics, LDA helps in understanding the structure of the data, which can be applied in various fields such as content recommendation, trend analysis, and document classification. The study practical side using Python and Gensim demonstrates how to implement LDA and interpret the results, showcasing its effectiveness in extracting meaningful insights from text data.

Study discovered many challenges faced Arabic national libraries in developing digital identities, such as the need for strategic alignment among their vision and mission statements and their digital presence.

In conclusion, the use of content analysis ML algorithms to create a digital identity for Arabic National Libraries is a powerful approach that aligns their digital presence with their core mission and values and support their roles in preservation, accessibility, cultural promotion, global recognition, innovation, and education by enhance their impact and relevance in the digital age. This method not only ensures a cohesive digital identity but also supports Arabic national libraries efforts to marketing their efforts for preserve and promote cultural heritage, contributing to the development of global civilization in digital World.

The other Arabic national libraries can leverage the experiences and best practices of leading institutions to enhance their digital presence. By adopting similar strategies, and they can make their digital identity more robust and engaging for a global audience.

For the case study libraries, developing multilingual digital platforms it is important aspect to cater to both Arabic and international audiences, and for enhancing accessibility and user engagement, and provide more collaboration with international digital library initiatives that can provide additional resources and expertise, facilitating the development of robust digital platforms and services.

Finally, Evaluating the accuracy of the suggested content analysis model can be done by using metrics such as precision, recall, F1-score, and confusion matrix, on the other hand the effectiveness of digital identities can be evaluated by Content Metrics such as website traffic, social media engagement, user feedback, and online resource usage to provide insights into the success of the digital strategies and areas for improvement.

Several future studies can be suggested to further explore and enhance using AI to create and supporting the digital identity for Arabic national libraries such as: Conduct a study to evaluate the impact of a strong digital identity on user engagement and satisfaction in national libraries. Explore the integration of multimodal data sources, such as text, images, and videos, to create a more comprehensive digital identity for national libraries. Explore the role of social media in building and promoting the digital identity of national libraries.

References:

1. Abdullah, Z., Anumudu, C. E., & Raza, S. H. (2022). Examining the digital organizational identity through content analysis of missions and vision statements of Malaysian and Singaporean SME company websites. *The Bottom Line, 35*(2/3), 137-158. https://doi.org/10.1108/BL-12-2021-0108

2. Abu Al-Khair, I. H., & Mohammed, M. I. H. (2006). *The status of Arab national libraries on the internet: An evaluative study*. Journal of Arts and Humanities [Minya], (60), 427-503.

https://www.researchgate.net/publication/330656052_mwaq_almktbat_alwtnyt_alrbyt_l
y_alantrnt_drast_tqyymyt

3. Allison, A., Currall, J., Moss, M., & Stuart, S. (2005). Digital identity matters. Journal
of the American Society for Information Science and Technology, 56(4), 364-372.
http://dx.doi.org/10.1002/asi.20112

4. Armann-Keown, V., & Patterson, L. (2020). Content analysis in library and
information research: An analysis of trends. Library & Information Science Research, 42,
101048. https://doi.org/10.1016/j.lisr.2020.101048

5. Atick, J. (2016). Digital identity: the essential guide. In ID4Africa Identity Forum
(Vol.   2016,   pp.   1-3).
https://www.id4africa.com/main/files/Digital_Identity_The_Essential_Guide.pdf

6. Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR
corporate disclosure. International Journal of Digital Accounting Research, 17.
https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2c4e6aff81a8357794
15af37551b12c3b7a50f19

7. Bos, W. & Tarnai, C. (1999). Content analysis in empirical social research.
International   Journal   of   Educational   Research,   31(8),   659-671.
https://www.doi.org/10.1016/S0883-0355(99)00032-4

8. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative
Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

9. British Library Labs. (2013). *Experiment with British Library's digital collections and
data*. Retrieved from https://labs.biblios.tech

10. British Library. (2020, August 25). *Supporting public libraries through a national
digital presence*. Living Knowledge Blog. Retrieved June 29, 2024, from
https://blogs.bl.uk/living-knowledge/2020/08/supporting-public-libraries-through-a-
national-digital-presence.html

11. Cameron,   K.   (2005).   The   Laws   of   Identity.   Microsoft   Corporation.
https://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf

12. Cavoukian, A. (2013). Privacy by Design: The 7 Foundational Principles. Information
and   Privacy   Commissioner   of   Ontario,   Canada.
https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf

13. Chaffey, D., & Ellis-Chadwick, F. (2020). Developing a digital marketing strategy:
Key   success   factors.   *Journal   of   Business   Strategy,   41*(1),   24-31.
https://doi.org/10.1108/JBS-10-2019-0190

14. Chang, C.H., & Hedberg, J.G. (2007). Digital Libraries Creating Environmental
Identity Through Solving Geographical Problems. International Research in
Geographical and Environmental Education, 16, 58 - 72. https://doi.org/10.2167/irg206.0

15. Chen T-H, Thomas SW, Hassan AE (2016) A survey on the use of topic models when
mining   software   repositories.   Empir   Softw   Eng   21(5):1843–1919.
https://doi.org/10.1007/s10664-015-9402-8

16. Chi, D.T. (2024). Role and functions of libraries in the information society. Scientific
and Technical Libraries. https://doi.org/10.33186/1027-3689-2024-4-56-67

17. Crowston, K., Allen, E., & Heckman, R. (2012). Using natural language processing
technology for qualitative data analysis. International Journal of Social Research
Methodology, 15, 523 - 543. https://doi.org/10.1080/13645579.2011.625764

18. CUMSPH. Columbia University Mailman School of Public Health. (2024). *Content
analysis*.   Columbia   University.   Retrieved   June   30,   2024,   from

https://www.publichealth.columbia.edu/research/population-health-methods/content-analysis

19. Damyanov, M. (2023, March 20). What is content analysis? Uses, types & advantages. Dovetail. https://dovetail.com/research/content-analysis/

20. Dancs, S. (2018). Information seeking and/or identity seeking libraries as sources of cultural identity. Library Management, 39, 12-20. https://doi.org/10.1108/LM-03-2017-0034

21. Delve, Ho, L., & Limpaecher, A. (2023c, February 15). *Content Analysis vs Thematic Analysis: What's the Difference?* https://delvetool.com/blog/content-analysis-vs-thematic-analysis

22. Drisko, J. W., & Maschi, T. (2016). Content analysis. Oxford University Press, USA. https://doi.org/10.1093/acprof:oso/9780190215491.001.0001

23. El Yazidi, R. (2024). Exploring the components of digital identity on social networks sites: Identifier, post, profile photo, and selfie. *European Scientific Journal, ESJ, 20*(1), 1. https://doi.org/10.19044/esj.2024.v20n1p1

24. El-Shawish, A. M. A. (2024). The digital identity of academics in the field of libraries and information in Egypt: An analytical study (Doctoral dissertation, Al-Azhar University, Faculty of Arabic Language in Menoufia, Department of Libraries and Information). Supervised by M. F. Abdel-Hady & F. M. Rabah.

25. ENISA. (2023). Digital identity standards: Analysis of standardization requirements in support of cybersecurity policy. ENISA. https://www.enisa.europa.eu/publications/digital-identity-standards .

26. European Commission. (2020). Self-sovereign identity and the eIDAS Regulation: A legal report (Final Report). Joinup. https://joinup.ec.europa.eu/sites/default/files/document/2020-04/SSI_eIDAS_legal_report_final_0.pdf

27. European Union. (1999). Council Directive 1999/93/EC of 13 December 1999 on a Community framework for electronic signatures. Official Journal of the European Communities. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31999L0093

28. European Union. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union. https://gdpr-info.eu/

29. Financial Action Task Force (FATF). (2020). *Guidance on digital identity: Appendice D*. Retrieved from https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Guidance-on-Digital-Identity-Appendice-D.pdf

30. Forman, J., & Damschroder, L. (2007). Qualitative content analysis. In Empirical methods for bioethics: A primer (pp. 39-62). Emerald Group Publishing Limited. https://www.emerald.com/insight/content/doi/10.1016/S1479-3709(07)11003-7/full/html

31. Fussell, R., Mazrui, A., & Holmes, N. G. (2022). Machine learning for automated content analysis: Characteristics of training data impact reliability. *Physics Education Research Conference 2022 Proceedings*. https://doi.org/10.1119/perc.2022.pr

32. GAO. U.S. General Accounting Office (1996). Content Analysis: A Methodology for Structuring and Analyzing Written Material. GAO/PEMD-10.3.1. Washington, D.C. https://www.gao.gov/assets/pemd-10.3.1.pdf

33. Gelb, A., & Clark, J. (2013). Identification for Development: The Biometrics Revolution. Center for Global Development. http://dx.doi.org/10.2139/ssrn.2226594

34. Grassi, P. A., Garcia, M. E. and Fenton, J. L. (2017), Digital Identity Guidelines, National Institute of Standards and Technology (NIST) Special Publication 800-63-3, NIST, Gaithersburg, MD. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-3.pdf

35. Harwood, T. G., & Garry, T. (2003). An overview of content analysis. The marketing review, 3(4), 479-498. https://moscow.sci-hub.se/4530/56fd933f632098115729205922edb6eb/harwood2003.pdf

36. Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online learning for latent Dirichlet allocation. *Neural Information Processing Systems (NIPS)*. https://doi.org/10.5555/2997189.2997274

37. IDunion. (2024). Digital identities for organizations (Version 1.1). Retrieved from https://idunion.org/wp-content/uploads/2024/03/IDunion_Digital-Identities-for-Organizations_v1.1.pdf

38. ISO, (2019). *ISO/IEC 24760-1:2019 Information technology — Security techniques — A framework for identity management — Part 1: Terminology and concepts* (2nd ed.). Retrieved from https://www.iso.org/obp/ui/en/#iso:std:iso-iec:24760:-1:ed-2:v1:en

39. Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

40. Joffe, H. (2011). Thematic analysis. In D. Harper & A. R. Thompson (Eds.), Qualitative methods in mental health and psychotherapy: A guide for students and practitioners (pp. 209–223). Chichester, UK: Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119973249

41. Krippendorff, K. (1980). Content Analysis: An Introduction to Its Methodology. Newbury Park, CA: Sage. https://www.daneshnamehicsa.ir/userfiles/files/1/9-%20Content%20Analysis_%20An%20Introduction%20to%20Its%20Methodology.pdf

42. Krippendorff, K. (2018). Content analysis: An introduction to its methodology. Newbury Park, CA: Sage publications.

43. Laurent, M., & Bouzefrane, S. (2015). Digital identity management. Elsevier. https://shop.elsevier.com/books/digital-identity-management/laurent/978-1-78548-004-1

44. Library of Congress. (2023). *Digital Strategy*. Retrieved from https://www.loc.gov

45. Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial intelligence review, 1-57. https://doi.org/10.1007/s10462-021-09973-3

46. Lyon, D. (2009). Identifying Citizens: ID Cards as Surveillance. Polity Press. https://www.surveillance-studies.ca/publications/identifying-citizens-id-cards-as-surveillance

47. M.B. Miles, A.M. Huberman. 1994. "Qualitative data analysis: A sourcebook" https://www.metodos.work/wp-content/uploads/2024/01/Qualitative-Data-Analysis.pdf

48. Meriam, M. (2023). *National libraries and their role in achieving sustainable development goals*. University of Guelma. Retrieved June 29, 2024, from https://tinyurl.com/4x2bbskm

49. Nastase, V., Koeszegi, S., & Szpakowicz, S. (2007). Content analysis through the machine learning mill. Group Decision and Negotiation, 16(4), 335-346. https://doi.org/10.1007/s10726-006-9053-7

50. National Library of Australia. (2023). *Digital strategy*. Retrieved June 29, 2024, from https://www.nla.gov.au/about-us/corporate-documents/corporate-strategies/digital-strategy

51. Neuendorf, K. A. (2017). The content analysis guidebook (2nd Ed.). Thousand Oaks, CA: Sage. https://www.daneshnamehicsa.ir/userfiles/files/1/9-%20The%20Content%20Analysis%20Guidebook.pdf

52. New-Zealand government. (2018) Digital Government. *Identification management standards* https://www.digital.govt.nz/standards-and-guidance/identity/identification-management/identification-management-standards/

53. Paperpile. (2024). How to do a thematic analysis [6 steps]. Retrieved June 30, 2024, from https://paperpile.com/g/thematic-analysis/

54. Patel, F. N., & Soni, N. R. (2012). Text mining: A Brief survey. International Journal of Advanced Computer Research, 2(4), 243. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ef81607cdbd353f9f3ebb8f16fd15b527b61f883

55. Pirnau, M.; Botezatu, M.A.; Priescu, I.; Hosszu, A.; Tabusca, A.; Coculescu, C.; Oncioiu, I. (2024). Content Analysis Using Specific Natural Language Processing Methods for Big Data. Electronics, 13, 584. https://doi.org/10.3390/electronics13030584

56. Prasad, B. D. (2008). Content analysis. Research methods for social work, 5(le20). In Lal Das, D.K and Bhaskaran, V (eds.). (2008) Research methods for Social Work, New in Delhi: Rawat, pp.173-193 http://dx.doi.org/10.13140/RG.2.1.1748.1448

57. Řehůřek, R. (n.d.). *Gensim: Topic modelling for humans*. Retrieved June 29, 2024, from https://radimrehurek.com/gensim/

58. Riffe, D., Lacy, S. & Fico, F. (2014). Analyzing Media Messages. Using Quantitative Content Analysis in Research (3rd Ed.). New York: Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9780429464287/analyzing-media-messages-daniel-riffe-stephen-lacy-frederick-fico-brendan-watson

59. Rothfus, M., & Helwig, M. (2019). Supporting digital identities at academic libraries. https://dalspace.library.dal.ca/bitstream/handle/10222/75725/Supporting%20Digital%20Identities%20APLA%202019_Rothfus_Helwig.pdf?sequence=1&isAllowed=y

60. Rubin, R. E., & Rubin, R. G. (2020). Foundations of library and information science. American Library Association. https://alastore.ala.org/flis5

61. Rundle, M., Blakley, B., Broberg, J., Nadalin, A., Olds, D., Ruddy, M., . . .Trevithick, P. (2007). At a crossroads: "personhood" and digital identity in the information society. STI Working Paper 2007/07. Organization for Economic Co-operation and Development. http://www.oecd.org/sti/working-papers

62. Scharkow, M. (2013). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity, 47*(2), 761–773. https://doi.org/10.1007/s11135-011-9545-7

63. SCORE. (2024). *How to set up an online presence for your business*. Retrieved June 29, 2024, from https://www.score.org/resource/blog-post/how-set-online-presence-your-business

64. Sedlmeir, J., Smethurst, R., Rieger, A., & Fridgen, G. (2021). Digital Identities and Verifiable Credentials. Business & Information Systems Engineering, 63, 603 - 613. https://link.springer.com/content/pdf/10.1007/s12599-021-00722-y.pdf

65. Sinha, M. K. (2022). Web directories: A searching tool. ResearchGate. https://www.researchgate.net/publication/362482640_Web_Directories_A_Searching_Tool

66. Škurtan, M. (2020). Digital Identity. Serif Affinity Publisher. Retrieved from https://joinup.ec.europa.eu/sites/default/files/document/2020-04/SSI_eIDAS_legal_report_final_0.pdf

67. Stemler, S. (2000). An overview of content analysis. Practical assessment, research, and evaluation, 7(1). https://typeset.io/papers/an-overview-of-content-analysis-53zvb3be0k

68. Sullivan, C. (2011). Digital identity: An emergent legal concept (p. 178). University of Adelaide Press. http://library.oapen.org/handle/20.500.12657/33171

69. Sullivan, C., & Burger, E. (2019). Blockchain, Digital Identity, E-government. In H. Treiblmaier, & R. Beck (Eds.), Business Transformation through Blockchain (Vol. II, pp. 233-258). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-99058-3_9

70. Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and sentiment analysis: A review in competitive research. *Computers, 12*(2), 37. https://doi.org/10.3390/computers12020037

71. Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences, 13*(7), 4550. https://doi.org/10.3390/app13074550

72. Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. International Journal of Computer Science & Communication Networks, 5(1), 7-16. https://doi.org/10.1016/j.eswa.2020.113851

73. Vitiello, G. (2001). National libraries: the concept and the practice, 1700–2000. Alexandria, 13(3), 139-151. https://doi.org/10.1177/095574900101300302

74. Wahdan, D. M. R. (2020). Electronic Presence of the Arabic National Libraries: Webometrics study. International Journal of Library and Information Sciences, 7(3), 285-304. https://journals.ekb.eg/article_121368_5f46a720c910b211ea3c28a825efe3a3.pdf

75. Weber, R. P. (1990). Basic Content Analysis, 2nd ed. Newbury Park, CA. https://www.daneshnamehicsa.ir/userfiles/files/1/9-%20Basic%20Content%20Analysis,%201990,%20Sage%20Publications.pdf

76. Weber, R. P. (1994). Basic Content Analysis. In Lewis-Beck, Michael S. (comp.): Research Practice (pp. 251-338). Thousand Oaks, CA: Sage.

77. Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). Text mining: predictive methods for analyzing unstructured information. Springer Science & Business Media. http://www.ttcenter.ir/ArticleFiles/ENARTICLE/3771.pdf

78. White, M. D. & Marsh, E. E. (2006). Content analysis: a flexible methodology. Library Trends, 55(1), 22-45. https://doi.org/10.1353/lib.2006.0053

79. Windley, P. J. (2005). Digital Identity. O'Reilly Media, Inc.

80. World Bank. (2018). *ID4D technical standards for digital identity*. Retrieved from https://thedocs.worldbank.org/en/doc/579151515518705630-0190022018/original/ID4DTechnicalStandardsforDigitalIdentity.pdf

81. World Bank. (2018). *Technology landscape for digital identification*. Retrieved June 29, 2024, from https://documents1.worldbank.org/curated/en/199411519691370495/Technology-Landscape-for-Digital-Identification.pdf

82. Zamith, R. & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What computational social science means for traditional modes of media analysis. The ANNALS of the American Academy of Political and Social Science, 659(1), 307-318. https://www.doi.org/10.1177/0002716215570576

83. Zyskind, G., & Nathan, O. (2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. In 2015 IEEE Security and Privacy Workshops (pp. 180-184). IEEE. https://homepage.cs.uiowa.edu/~ghosh/blockchain.pdf

---

[1] The study follows APA style for Citation.

[2] https://www.nla.ae/en/

[3] https://kfnl.gov.sa/Ar/About/Pages/VisionKfnl.aspx

[4] https://qnl.qa/en/about/vision-and-mission

[5] https://www.nlk.gov.kw/MissionVision.aspx

[6] https://www.icc.gov.bh/who%20we%20are/AbouttheCentre/

[7] https://www.nlp.ps/en/node/5