

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
**ECOLE SUPERIEURE EN INFORMATIQUE 08 MAI 1945 SIDI BEL
ABBES**



Thesis

In partial fulfillment of the requirements for the **Engineering's** degree

Field: **Computer Science**

Field : **Ingénierie des Systèmes Informatiques (ISI)**

Topic

**Development of a web interface for the analysis and search
of French literary data lakes for a LIFRANUM project.**

Presented by :

M^{lle} Djouad Kawther

Soutenu le : xx/09/2023

Devant le jury composé de:

Présidente:	<i>M^{me}</i> ...	Maître de conférences A
Examinatrice:	<i>M^{me}</i> ...	Maître de conférences A
Encadrant:	<i>M^r</i> MAHAMMED Nadir	Maître de conférences A
Co-encadrant:	<i>M^r</i> FAHSI Mahmoud	Maître de conférences A
Co-encadrant:	<i>M^r</i> VELCIN Julien	Professeur B

Academic year 2022/2023

Acknowledgements

I am grateful to Almighty Allah for giving me the strength and ability to fulfill this work .

First ,i would like to express our deep gratitude to our teacher and supervising Mr Nadir MAHAMMED for his follow-up and his enormous support throughout the period of preparation of the dissertation. I hold to also thank my co-supervisor Mr Julien VELCIN for the time he devoted with interest and understanding, and for all the encouragement and opportunities he invited me to, only when I thought I would never have them, Not forgetting to Thank Mr. Mahmoud FAHSI for his great collaboration.

Secondly, I would like to extend my sincere thanks to the members of the jury for agreeing to examine, evaluate, and judge my work.

Thirdly, I would like to express my sincere gratitude to the ERIC and LIFRANUM teams for granting me the opportunity to collaborate with them on such an important project.

Last but not least, I must acknowledge our institution, ESI SBA, for providing us with the necessary environment and knowledge that enabled us to successfully undertake and complete such significant projects.

I will not let this opportunity pass, without thanking my parents and all my family who , through their prayers and encouragement, I was able to overcome all the obstacles. I would like to thank everyone who participated directly or indirectly in the execution of this work.

Dedication

I dedicate this modest work to those closest to my heart:

To my parents who accompanied me during the toughest moments of this long journey of my education , who shared with me all the emotional moments and who always supported and encouraged me . I wish to show my great gratification which will never be enough and that I hope to make them proud of this work.

To my sisters AHLAM and MERIEM, To my nieces Taouba and Line, And to my gardening angel Brother SALAH as a testimony of fraternity with my wishes for happiness, health and success.

To my grand-father for his love and prayers. To my Childhood friend Asma and her family, To my uncle Boulaaras and his family who both made a second home for me for 5 years. And to all the members of my family and all my friends who were with me during this great journey,

Abstract

In the age of information abundance, data lakes have emerged as a foundational element in the management of big data. These centralized repositories offer organizations the ability to efficiently store and analyze vast and diverse datasets from a multitude of sources, including social media, IoT devices, and enterprise systems. Data lakes provide a scalable and cost-effective solution, allowing data to be ingested in real-time or batches, stored in distributed file systems or cloud platforms, and processed using powerful frameworks like Apache Hadoop and Apache Spark. Effective data lake management entails meticulous data organization, metadata management, and data governance to ensure data accuracy, discoverability, and usability, empowering businesses to extract valuable insights and make data-driven decisions in an increasingly data-driven world.

Keywords : Big Data, Data Lakes, NLP, Information retrieval, Mongoddb, Dash library.