Movie Recommendation System Based on Semantic Web and Machine Learning

Venu V[1], Dr. T Vijaya Kumar[2]

[1] *Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India*

[2] *Professor & Head, Department of MCA, Bangalore Institute of Technology, Karnataka, India*

---***---

Abstract - It is difficult to acquire, from these massive fragmentary data, valuable design knowledge for designers in conceptual product design.
The paper introduces a new approach to design knowledge acquisition through the integration of deep learning with knowledge graph techniques. In particular, through the knowledge extraction model, design-related entities and their relations can be extracted from fragmentary data, and then the knowledge graph is constructed to support design knowledge acquisition for conceptual product design. Apart from that, the knowledge extraction model proposes ALBERT to resolve the problems of memory limitation and communication overhead in the entity extraction module, and applies multi-granularity information to segment errors and polysemy ambiguity in the relation extraction module. Experimental comparison verified that the proposed knowledge extraction model was effective and accurate. In this case study, it could prove that the knowledge graph construction is feasible with real fragmentary porcelain data. More than this, it has shown capability in providing designers with connected and visualized design knowledge.

Key Words: Knowledge Extraction, Knowledge Graph, Deep Learning, and Knowledge Acquisition.

1.INTRODUCTION

For a system to be viewed as "intelligent," certain properties should be fulfilled: it has to adapt and react unknown, and it needs to have some kind of understanding of the world. Wherever we go, one way or another, intelligence is being applied—either through Artificial Intelligence or some other techniques.
The combination of Semantic Web and Machine Learning gives more impact on Intelligence, where it will act subjected to constant refinement over time while it obtains access to new information. Semantic Web or Web3.0 is all about a web of data rather than documents.
Structurally, this linking of interrelated data on the web can also be referred to as Linked Data.
Basically, the Semantic Web is for enabling Internet data to be interpretable to humans and machines. This would foster automation of lots of work on the web with the help of online agents——in the sense, Software. If we go Semantic Web, that is to present knowledge about our data, to allow data integration, and to bring intelligence to our system.
After going through some research papers, we got more knowledge about the Semantic Web Technologies. All concentration of the semantic web is mainly into three areas: OWL, RDF, and Ontology. Ontology is a model of knowledge defining a certain set of concepts and relationships between those concepts within a particular domain. This mainly supports the automated reasoning and inference of data based on the prescribed technique using logical rules for such a purpose. It provides knowledge sharing and reuse among people or software agents. One of the Ontology techniques is the Knowledge Graph. Knowledge Graph mainly describes entities of the real world and their interrelations, which is organized in a graph.

It enables any arbitrary entities to potentially interrelate with each other and covers a wide range of topical domains when schema is defined for possible classes and relations of entities within the Knowledge Graph. The KG is one of the major application for recommend movies, Movie Recommendation System is one of the application to recommend best movies.

2. LITERATURE SURVEY

This research area of SWeML has gained a lot of traction in the past few years, as shown in a rapidly growing number of publications in different outlets, as well as SWeML techniques being employed to solve problems in various domains. At the same time, this growth poses two main challenges that threaten to hamper further development of the field.
First, keeping up with the main trends in the field has become unfeasible, not only because of its fast pace and a large volume of published papers but also because papers require understanding techniques from the two diverse research sub-areas of AI. In an attempt to address this challenge, several works aimed to provide overviews of SWeML Systems and related systems. However, reviewing those, we conclude that existing work either (1) focuses rather on a wider or related field or, on the contrary, (2) is scoped around a very specific sub-field of SWeML. Additinally, none of the reviewed surveys adopts a principled and reproducible methodology that would guarantee unbiased and representative data collection. We therefore conclude that there is a need for a survey that adopts a solid review methodology to complement current insights with evidence-based findings. The second challenge, which amplifies the first one, is the lack of a standardized way to report SWeML Systems that hampers understanding all key aspects of these systems. On the one hand, authors of SWeML Systems would benefit from a structured way to describe their system and its key characteristics. Readers, on the other hand, would benefit from a structured way of interpreting such systems. This would not only facilitate the understandability for those coming from other communities but also improve the comparability of different systems. An early work in this direction was proposed by Van Harmelen and ten Teije by

introducing patterns for representing hybrid AI systems in terms of their components and information flows with the aim to facilitate a more schematic representation of the system. Although these patterns were derived from a large number of papers, there is currently no insight into their adoption in the field (e.g.,about the completeness of the introduced system patterns) or their usage frequency. First, when the problem statements are defined, we identify scientific databases and search keywords that will be used to select primary studies. To query the selected scientific databases, a combination of two main keywords is used: knowledge graph and machine learning. The number of obtained articles is large, and not all of these studies are relevant for our review. Afterwards, we refined the query by restricting the search of these same keywords in title, abstract, and keywords, which gave a much smaller number of articles that look more relevant. After first analysis, we realize that the term Deep Learning is often used instead of Machine Learning even though it is a subset of this technique. As that, we included the deep learning keyword to our query, putting the same restrictions applied to the machine learning term. Also, some authors use directly the term neural network (in particular for the most recent articles) without explicitly mention the machine learning or deep learning terms. We added we therefore decided to add this keyword to our query, even if it concerns a minority of articles (10 to 15%) in each query. For the knowledge graph keyword, we decided to use only this term and not combine it with other keywords representing different semantic techniques, such as taxonomy, knowledge modeling, or ontology. In this review, we are mainly interested in the use of knowledge graphs, and the possibility to perform logical reasoning. Finally, we have also included in our query the artificial intelligence keyword. This allows to restrict obtained results to our research domain. Unlike previous keywords, the presence of this term is looked for anywhere in the article to be less restrictive. Based on our search, the selected keywords are: Knowledge Graph, Machine Learning, and Artificial Intelligence. Most of the added inputs in the studied systems were symbolic; therefore, a trend in exploiting symbolic representations to augment system functionalities can be outlined. Nevertheless, it is interesting to note how one fifth of the authors used additional non-symbolic inputs mainly to systems using Graph-based deep learning methods or when improving Generalization or Usability. Their presence all together is an interesting aspect because Explainability and Usability often co-occurred. objectives in many systems, driving their remarkable connection. Fur-thermore, we discovered a category of hybrid systems that provide the flexibility to consider either symbolic or non-symbolic inputs as optional. Despite Performance objectives dominating the research landscape, it exposes a wide variety of goals, and the growing count of "multiple-objective seekers." This trend mirrors the evolving needs within the AI field. It would underscore systems needing to be designed with a fuller range of associated requirements for effective performance and adaptation.[3]

3. EXISTING SYSTEM

The existing systems mainly focus on that means that when the user makes the query in the search, all relevant pages are drawn from the index and algorithms arrange the relevant pages in order of hierarchy into a result set.

That is, ranking of the most relevant results differs for each search engine's algorithms. Again it works only based on keyword matching: wherever the user-typed keyword is present, those the pages will be displayed. That will be resolved in the proposed system.

Related Survey Papers Focusing on the Combining of Knowledge Graph and Machine Learning

Title	Author	Technologies	Year	Published in	Ref
Combining ML & SW: Systematic Mapping Study	Anna Berite Luara Waltersdorfer	SW, ML, AI, XAI, Data Mining, Semantic Web mining, Knowledge Representation & Reasoning Knowledge Graph, Systematic Mapping Study.	July 2013	ACM Computing Surveys	[1]
Combing ML & ontology: A Systematic Literature Review	Saraha Ghidalia Christophe Nicolle Aurelie Bertax	ML, AI, Hybrid AI, Dl, NLP, Neuro-Symbolic Approaches.	Feb 2024	ACM Computing Surveys	[2]
Analysing the motivation behind SW, ML System Design	Artem Revenko	Neuro-Symbolic AI, ML, SWeML, KR			[3]
Semantic web technologies for explainable ML models : A literature review	Anne seeliger, Matthias Dfaff, Helmut Karctnar	Semantic Web Technologies, ML, XAI Black-box ML Models, Neural networks, KG, ontology, Taxonomy, MDP(Reinforcement).	Oct 2019		[4]
Knowledge Graph base Approach for predicting future research collaboration	Nikos Kanakaris, Nikolas Giarelias, Tlias Siachos	Python Programming, Tensor & sci hit-Learn Grah Theory, Graph-based text representation, NLP & KG representation	May 2021	https://www.mdpi.com/journal/entropy	[5]
Research on the application of semantic network in disease diagnosis prompts based on medical corpus	Yufeng li, Weimin Wong, Xu Yan, Min Gao, Ming Xuan Xiao	Disease-Symptom Semantic network, int Ontology.	Feb 2024	IJIRCST	[6]
A Hybrid semantic Knowledgebase ML Approach for opinion mining	Rowida Alfrjioni, Tah Osman, Georgina Cosma	NLP, SPARQL, Opinion Classification, Domain Feature extraction Sentiment extraction.	May 2019	www.elsiver.com/locate/datak	[7]
KG & DL based Pest detection & Identification System for fruit quality	Ding Ju Zha, LionZi Xie, BingXu Chen, JionBin Tan, RenFeng Deng	Knowledge graph, Raspberry PI, Integrated Pest Management, Agriculture Knowledge Map DL, Knowledge Fusion, Knowledge Storage.	Nov 2022	http://doi.org/10.1016/jiot.2022.100649	[8]

Fig 1 - Model Block Diagram

4. PROPOSED SYSTEM

The proposed system helps in Movie Recommendation System based on Combining of Knowledge Graph and Machine Learning is to find the synonyms of the user-typed keyword and fetch the movie wherever the relevant movie name is present in the current web pages that will be recommended.

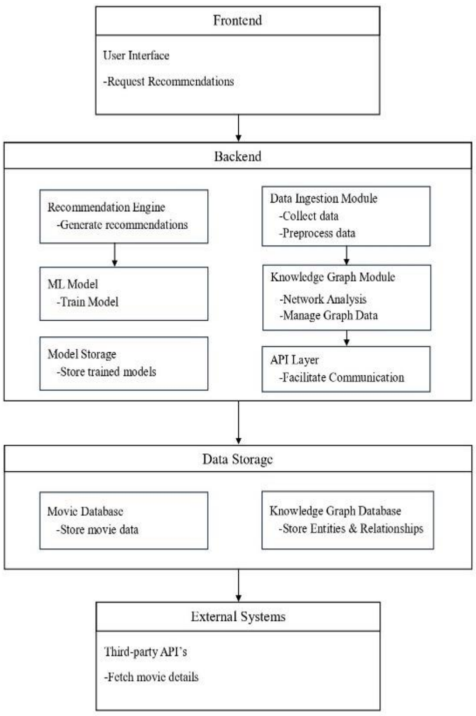

***Fig-1**: System Architecture*

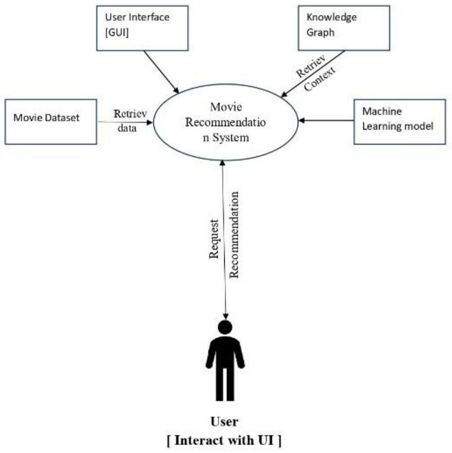

Fig-2: Use Case Diagram

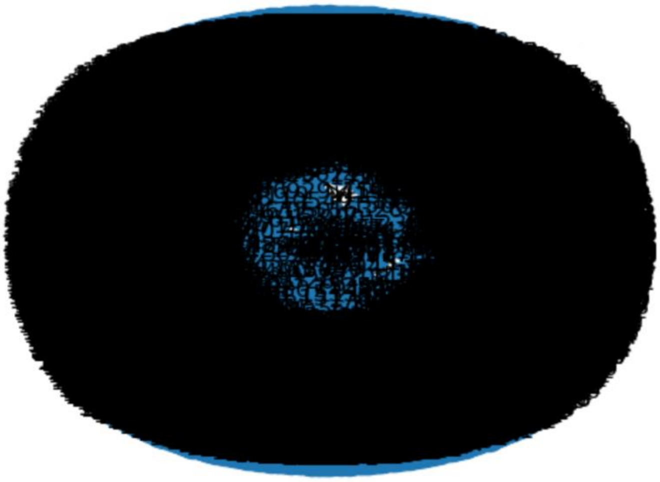

Fig-1: Constructed Knowledge Graph

5. IMPLEMENTATION

1.Data Collection: Data collection refers to the process of gathering and measuring information on variables of interest in an organized way. Gathering and compiling data are essential parts of research, analysis, and decision-making in many disciplines, directly or indirectly related to science, social sciences, business, and technology.

2.Dataset: df This will be the variable name that holds the result DataFrame. In Python, it's pretty standard to use df as shorthand for "DataFrame".

pd : This is an alias for the pandas library. You would generally put the import statement for the pandas library at the top of your script: import pandas as pd in Python. The as pd part creates an alias called pd for the pandas library so that when you want to use any functions or classes from the library, you can simply use pd.

read_csv—One of the pandas library functions that reads a Comma Separated Values, CSV, file into a DataFrame. The read_csv function represents a fairly robust but somewhat flexible way to import data from a text file into a Pandas DataFrame'.

3.Data Preparation: Data preparation, or preprocessing of data, is an exceedingly important stage in the pipeline of any analysis of data. It comprises translation of raw data into a form that will let it be analyzed and modeled. At this step, it will be ensured that prepared data are clean, organized, and relevant for any analysis to be performed efficiently and effectively.

4.Network Analysis:

Network analysis, sometimes referred to simply as network or graph theory, is a subfield of mathematics and computer science that deals with the study of structure, properties, and dynamics of networks. It provides insight into the heterogeneous relationships and interactions of nodes (usually entities) connected by edges in data science and social sciences.

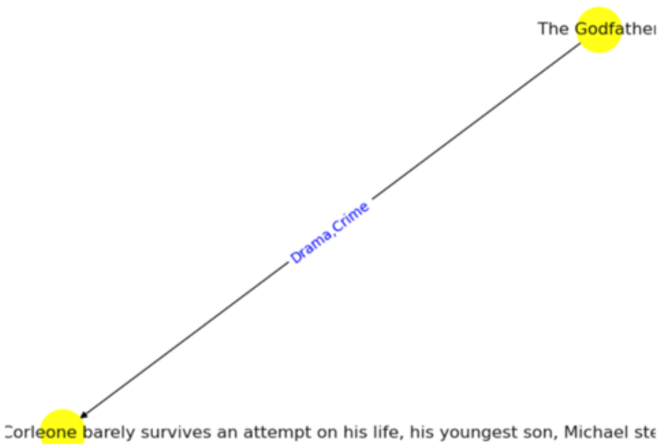

Fig-2: Sub Graph

5.Machine Learning Model: A model in machine learning is a mathematical representation or algorithmic structure that learns from data. It is designed to make predictions, classification, or decisions for input data without being explicitly programmed for specific tasks.

5. RESULTS

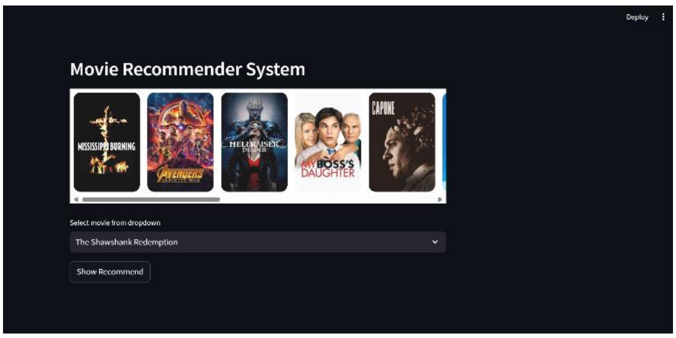

Fig-1: Home Page

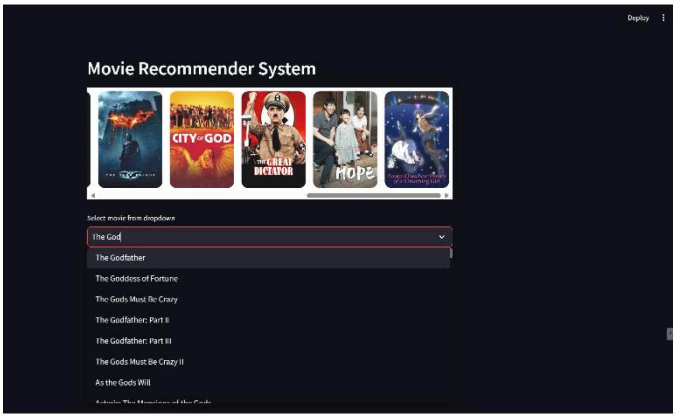

***Fig-2**: User Searching Page*

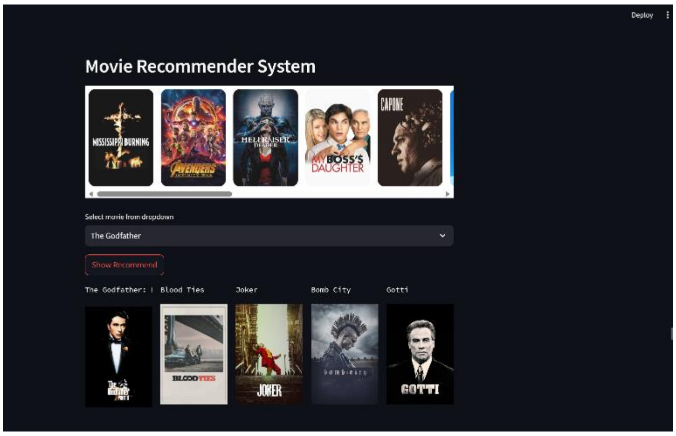

***Fig-3**: Movies Recommended Based on User Search*

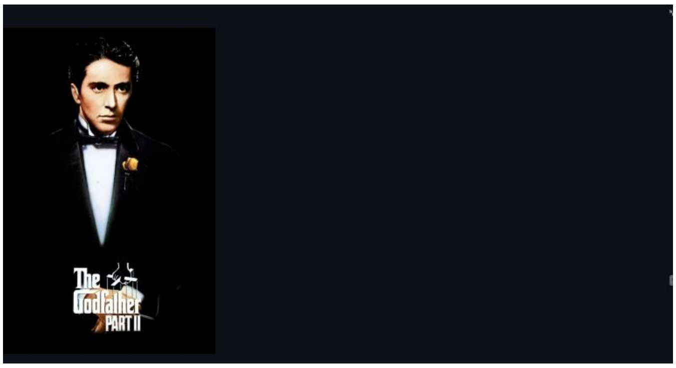

***Fig-4**: Review of Movie*

6. CONCLUSIONS

A movie recommendation system using knowledge graphs and machine learning would be a robust and sophisticated approach to personalized content delivery. In the proposed dual-method strategy, it will improve upon the recommendation quality by combining rich context from knowledge graphs with inductive power from machine learning algorithms.It not only improves user satisfaction but also maintains the elasticity and no-end point tendency of the system, able to evolve with new data and user preference.

7. FUTURE ENHANCEMENTS

Other improvements will involve increasing the datasets to be more diverse and optimizing the models to process in real-time. Integration into various other fields will make them much more useful. Semantic Web Technology is a phenomenal improvement of the representation, sharing, and usage of data on the web. It provides a framework that may tell machines what information means and thereby assists in the development of more intelligent, more efficient data processing for developing smarter applications and more interconnected web resources.

8. REFERENCES

1. ANNA BREIT, LAURA WALTERSDORFER, FAJAR J. EKAPUTRA, MARTA SABOU, ANDREAS EKELHART, ANDREEA IANA, ARTEM REVENKO, ANNETTE TEN TEIJE and FRANK VAN HARMELEN.
"Combining Machine Learning and SemanticWeb: A Systematic Mapping Study".

https://doi.org/10.1145/3586163

2. . Sarah Ghidalia, Ouassila Labbani Narsis, Aurélie Bertaux, Christophe Nicolle.
"COMBINING MACHINE LEARNING AND ONTOLOGY: A SYSTEMATIC LITERATURE REVIEW".

https://www.w3.org

3. Artem Revenko(Semantic Web Company).
"Supplementary Objectives: Analysing The Motivations Behind SemanticWeb Machine Learning System Design".

The First Austrian Symposium on AI, Robotics, and Vision (AIROV24)

4. Arne Seeliger, Helmut Krcmar, Matthias Pfaff.
"Semantic Web Technologies for Explainable Machine Learning Models: A Literature Review"

https://www.researchgate.net/publication/336578867

5. Nikos Kanakaris, Nikolaos Giarelis, Ilias Siachos and Nikos Karacapilidis.
"Shall I Work with Them? A Knowledge Graph-Based Approach for Predicting Future Research Collaborations"

https://doi.org/10.3390/e23060664

6. Yufeng Li, Weimin Wang, Xu Yan, Min Gao, and MingXuan Xiao.
"Research on the Application of Semantic Network in Disease Diagnosis Prompts Based on Medical Corpus". https://doi.org/10.55524/ijircst.2024.12.2.1

7. Rowida Alfrjani, Taha Osman ∗, Georgina Cosma.
"A Hybrid Semantic Knowledgebase-Machine Learning Approach for OpinionMining" www.elsevier.com/locate/datak

8. DingJu Zhu, LianZi Xie, BingXu Chen, JianBin Tan, RenFeng Deng et al.
"Knowledge graph and deep learning based pest detection and identification system for fruit quality" www.sciencedirect.com/journal/internet-of-things

9. Yuexin Huang, SuihuaiYu, JianjieChu1, Zhaojing Su, Yangfan Cong, HanyuWang and Hao Fan. "Combining Deep Learning with Knowledge Graph for Design Knowledge Acquisition in Conceptual ProductDesign"

10. Marko Horvat, Andrija Krtali´c, Amila Akagi´c and Igor Mekterovi´c.

"Ontology-Based Data Observatory for Formal Knowledge Representation of UXO Using Advanced Semantic Web Technologies"
https://doi.org/10.3390/electronics13050814

11. Huimin Luo, Weijie Yin, Jianlin Wang, Ge Zhang, Wenjuan Liang, Junwei Luo, and Chaokun Yan.
"Drug-drug interactions prediction based on deep learning and knowledge graph: A review"

http://creativecommons.org/licenses/by-nc-nd/4.0/

12. Youtube(https://www.youtube.com/watch?v=gtGKkBFuvlw&list=PLNXdQl4kBgzubTOfY5cbtxZCgg9UTe-uF&index=13)