

Knowledge Graph based Intelligent Conversational Agent for UK Immigration Case Work^{*}

Smriti Kotiyal¹, Dhavalkumar Thakker¹, and Yash Dubal²

¹ University of Bradford, Bradford UK

² A Y & J Solicitors, London UK

Abstract. Artificial Intelligence (AI) is reaching every domain relevant to human life. However, within a traditional domain like law, and specifically Immigration Law (IL), the reach of AI has been limited. The plethora of knowledge available within the UK IL archives, tacit knowledge from immigration solicitors and case workers and all other sources of information remains underutilised in building AI based systems. This paper presents early results from the first ever Innovate UK funded project on the use of AI in IL. The scope of the project is to design a decision support system (DSS) using Knowledge Graph (KG) technology for serving clients and preparing cases for immigration case workers. This paper presents the first substantial KG, capturing the knowledge from IL experts and archives, and the design and development of an AI based conversational agent that utilises the KG.

Keywords: Knowledge Graph · Conversational Agent · Immigration

1 Introduction

The UK visa immigration (UKVI) system is an intricate network of rules which is accessed for information by immigration case workers and applicants while making visa applications [1, 2]. This complexity of finding answers surfing through multiple IL archives can be a laborious job even for an expert. The applicants in turn pursue experts' advice for simple queries at costly rates. KG has not been efficiently utilised in this domain [3]. To address this challenge, A Y & J Solicitors³ and the University of Bradford⁴ have come together to work on a Knowledge Transfer Partnership (KTP) project funded by Innovate UK, to transform specialist knowledge and expertise into KGs to develop an intelligent conversational agent (ICA). The UK immigration information is available in a wide variety of sources like the UKVI website [4], UK IL archives [5], tacit knowledge from immigration solicitors and case workers, communication scripts between the experts and the clientele, and so on. With this project, we have

^{*} Supported by Innovate UK

³ <https://ayjsolicitors.com/>

⁴ <https://www.bradford.ac.uk/external/>

captured this knowledge into a KG and utilised it to build an AI system called LILA (Legal Immigration Artificial Intelligence Advice) which is available as a conversational agent. With the advent of LILA, we provide the clients a platform for initial customer interaction which would help the immigration experts in A Y & J with the increased influx of clients.

This paper provides an account of the design and implementation of the KG for ICA built with focus on the “Skilled Worker” (SW) visa category. It also provides the early results of the ICA, which understands the queries, reasons and traverses through the KG to provide consolidated and most appropriate answers, which are conventionally scattered through multiple sources of IL information. We tested our solution against a set of questions and achieved results with more than 80% accuracy. The scope of the project extends to building LILA as a DSS to help the immigration experts and case workers in the company with visa application preparation and client’s information collection.

2 Related Work

The growing amount of enriching information in the world makes it exceedingly essential for it to be more queryable to fetch relevant search results for the users. The development and use of question answering (QA) systems is ongoing for many years including domains’ specific ontology systems [6]. Few advances have been made in QA systems in the legal domain [7, 8] and problem-based reasoning [9]. In law, however, the potential of AI extends to deeper dimensions which are continually being explored. One such project was taken up by the researchers in the Oxford University which aimed to explore and implement the potential of AI in English Law domain [10]. Legal expert systems are a focus of research since 1980s and there is a significant need to develop legal DSSs to provide free legal services to the users [11]. We follow the same ambition of using AI to provide approachability to the non-lawyers with informed consent [12] using our domain-based KG. Our system is the first of its kind to be built for the UK’s Immigration law domain.

3 Knowledge Graph Development for Intelligent Conversational Agent

We have followed METHONTOLOGY methodology for developing the KG which is a well-structured way to build ontologies from scratch [13]. Initially, a workshop was held with the immigration solicitors and case workers in A Y & J Solicitors, to understand the process and conditions of a SW visa. Protégé [14] was used to build the KG model using the information retained from the workshop. Figure 1 shows a structural example of one of the eligibility requirements modelled in the graph for SW visa. The nodes or classes (circles) in the graph have multiple data and object properties to widen the search criteria. The figure shows the instances (diamonds) of the point-based system class as associated with the Eligibility criteria for a SW visa.

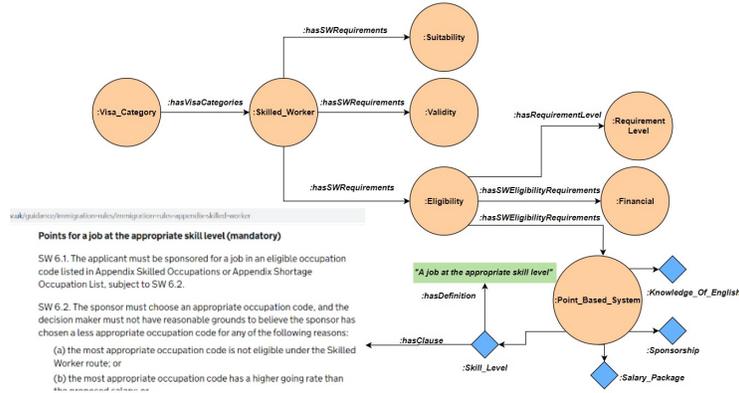


Fig. 1: KG Model : Example of eligibility requirement model for SW visa

subject	predicate	object	id	subject	predicate	object	id
sw_Eligibility_PBS_Skill_Level	hasPointsAwardOf	"20"^^xsd:integer	1	CSW6	hasRuleHeaderValue	"Points for a job at the appropriate skill level (mandatory)"	1
sw_Eligibility_PBS_Skill_Level	hasDefinition	"A job at the appropriate skill level"	2	CSW6	hasRuleValue	"Sw 6.1. The applicant must be sponsored for a job in an eligible occupation code listed in Appendix Skilled Occupations or Appendix Shortage Occupation List, subject to Sw 6.2."	2
sw_Eligibility_PBS_Skill_Level	hasRequirementValue	"RQF 3 or above"	3	CSW6	hasRuleValue	"Sw 6.2. The sponsor must choose an appropriate occupation code, and the decision maker must not have reasonable grounds to believe the sponsor has chosen a less appropriate occupation code for any of the following reasons: <p>(a) the most appropriate occupation code is not eligible under the Skilled Worker route; or</p> <p>(b) the most appropriate occupation code has a higher going rate than the proposed salary; or</p> <p>(c) the most appropriate occupation code is not a shortage occupation and the applicant is claiming points for a job in a shortage occupation; or</p> <p>(d) the most appropriate occupation code is not listed as 'eligible for PhD points' in Table 1 of Appendix Skilled Occupation List, and the applicant is claiming points for an educational qualification."</p>	3
sw_Eligibility_PBS_Skill_Level	rdf:type	"skill level"	4	CSW6	hasRuleValue	"Sw 6.3. To support the assessment in SW 6.2, the decision maker may, in particular, consider: <p>(a) the most appropriate occupation code is not eligible under the Skilled Worker route; or</p> <p>(b) the most appropriate occupation code has a higher going rate than the proposed salary; or</p> <p>(c) the most appropriate occupation code is not a shortage occupation and the applicant is claiming points for a job in a shortage occupation; or</p> <p>(d) the most appropriate occupation code is not listed as 'eligible for PhD points' in Table 1 of Appendix Skilled Occupation List, and the applicant is claiming points for an educational qualification."</p>	4
sw_Eligibility_PBS_Skill_Level	hasComment	"valid skill level"	5				
sw_Eligibility_PBS_Skill_Level	hasClause	CSW6	6				
sw_Eligibility_PBS_Skill_Level	hasSOCCodeRequirement	Eligible_SOC_Codes	7				
sw_Eligibility_PBS_Skill_Level	hasRequirementLevel	Mandatory	8				
sw_Eligibility_PBS_Skill_Level	hasSOCCodeRequirement	Shortage_Occupation_Eligible_SOC_Codes	9				
sw_Eligibility_PBS_Skill_Level	hasDetailRequirements	SW_ATAS	10				
sw_Eligibility_PBS_Skill_Level	rdf:type	sw_Eligibility_PBS	11				
sw_Eligibility_PBS_Skill_Level	hasDetailRequirements	sw_Eligibility_PBS_Sponsorship	12				
sw_Eligibility_PBS_Skill_Level	hasDetailRequirements	sw_Job_Demonstrates	13				
sw_Eligibility_PBS_Skill_Level	rdf:type	swNamedIndividual	14				

(a) Skill Level instance

(b) Skill Level clause CSW6 instance

Fig. 2: Instances as saved in the KG

One of the challenges of answering legal queries is to provide appropriate proofs to support the answers in the form of Law Acts and Clauses. Considering this, a semi-automated approach was adopted to acquire knowledge into the KG by developing a *Python* solution. It uses *Beautiful Soup* and *Owlready2* packages to parse the IL archives and save the fetched information into the instances of a KG "Clause" class respectively. These instances would point to the relevant entities of the KG via the object property *hasClause*. In Figure 1, the *Skill_Level* instance has two properties: *hasDefinition* that defines the instance, and *hasClause* that connects it to the exact clause associated with "Points for a job at the appropriate skill level(mandatory)" clause of the UK IL archives. Hence, *Skill_Level* instance becomes the data centre for all the information specific to skill level of the offered job. The information is saved in the KG as shown in Figure 2.

4 Development of Intelligent Conversational Agent

Architecture To utilise the developed ontology in section 3 into an ICA, a Python solution has been developed which connects the user interface with a text processing toolkit called GATE [15]. The GATE developer performs information extraction from the queries’ corpus using its processing resources (PR) which include tokenizer, named entity and parts of speech recognizer. The ontology is loaded into GATE using its *OntoGazetteer* PR which maps the corpus to the ontology classes to generate *Lookup* type annotations. A question can be asked in multiple ways. For example, both the questions “*My passport is expiring, is this a problem?*” and “*Will I have any problem if my passport expires soon?*” have the same meaning. However, our solution classifies these questions into 2 different question templates (QT) according to the way they are constructed. We use GATE’S JAPE grammars to classify queries into relevant QTs. JAPE has a set of patterns/rules which may or may not contain regular expressions to match the different QTs and log it into the annotations. All these annotations, consisting of the ontology matched entities’ URIs, token types, QT and their offsets, are sent back to the solution as output. Python algorithms have been developed to process each QT which uses the *QT* class output from GATE to select the correct algorithm and *annotations* output to formulate a SPARQL query against the selected algorithm. This is then used to fetch answers from the KG which are further structured into a human readable format and sent to the user interface as shown in figure 3.

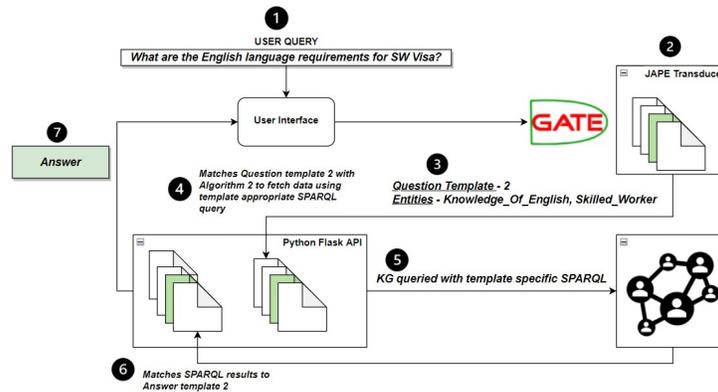


Fig. 3: ICA System Architecture

Question Templates We collected a set of 239 frequently asked questions from online resources to analyse their types and ways they are structured. We, therefore, classified them into 6 QTs as shown below to be used as explained above. In section 5, we will discuss the implementation of algorithm for QT1.

- Type QT1 - Stated Fact. Followed by a question?
- Type QT2 - How_____?
- Type QT3 - Can/Will/Would_____?
- Type QT4 - Do/Does/Should_____?
- Type QT5 - What_____?
- Type QT6 - Which/Who/When/Where/Why_____?

5 Implementation

Following the system architecture in figure 3, we have developed a Python Flask solution which provides a user interface as per figure 5. The solution communicates with GATE using Python *GateNLP* package and extracts relevant entities and relations. We have developed an algorithm for processing QT1 which uses SPARQL to find a relationship between the fetched entities which can be classes or instances, making a subgraph from the KG. Figure 4 gives the working iterations of the algorithm with an example. The found relations are used to fetch their definitions or *isDefinedBy* annotations from the KG to form an answer.

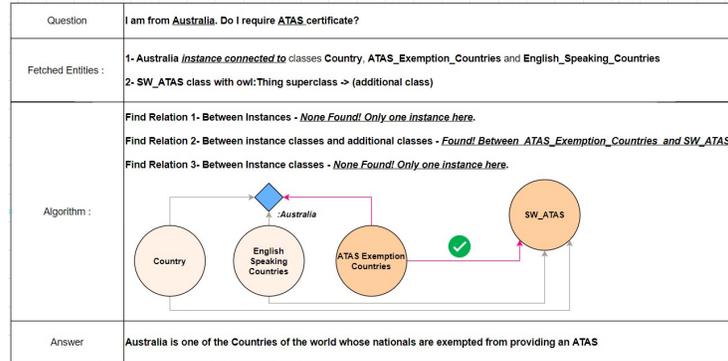


Fig. 4: Algorithm for question template 1

The KG plays an essential role in generating annotations from the user input. To facilitate better matching, Python *NLTK WordNet* has been used to extract synonyms, hyponyms and hypernyms of the KG nodes' names and added as respective class annotations using Protégé. Another set of word dictionary has also been created by measuring the average of WuPalmer – Wordnet similarity [16] between the class name and words *immigration* and *law*. This is to filter out many irrelevant words to make the process more efficient. To examine their utility in information extraction from KG, 6 different sets of word dictionaries have been created as under.

- Set 1 - With all synonyms, hyponyms and hypernyms

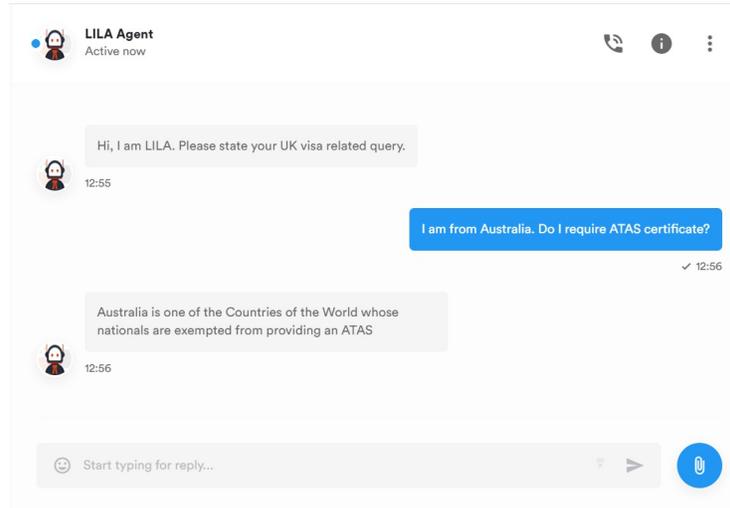


Fig. 5: User Interface of ICA

- Set 2 - With only synonyms and hyponyms
- Set 3 - With only synonyms
- Set 4 - With all synonyms, hyponyms and hypernyms with wup_similarity
- Set 5 - With all synonyms and hyponyms with wup_similarity
- Set 6 - With only synonyms with wup_similarity

Tables 1 gives the test questions, expected answers and evaluation results across the above 6 sets with the KG. According to the results, set 1 gives the most inaccurate answers. It is seen to match irrelevant classes from the ontology, for example, for Q1, it matches “am” keyword with *Eligibility* class taking reference from the word “be” added to the ontology for the class. Set 4 provides the most accurate results in this test by appropriately annotating the relevant tokens. It also provides correct results for Q5, by annotating the token “denial” with *Grounds_of_Refusal* class in the ontology. Ideally, set 1 makes this annotation too but fails while computing through the developed algorithm to fetch the correct answer. In this way, the queries are understood and answered by the system for QT1 which works well with more than 80% accuracy.

6 Conclusion and Future Work

The presented system offers answers to basic SW visa related queries which can be straightforward or complex. This is the very first version of the ICA as part of our project due for production deployment. The system performance is as expected at this point and the integration of algorithms for rest of the QTs are in progress. We are also aware of how the information in the IL archives captured in the KG can change overtime. Hence, we are also working on making

the KG semi-dynamic where it detects any changes in the UK IL and can allow the experts to insert them with sanity and consistency checks. The next step is to improve the system upon feedback from the users and to upgrade it with an ability to ask follow-up questions and collect information for visa application preparation.

Table 1: QT1 type questions, expected answers and evaluation results

Questions	Expected Answers	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
I am from Australia. Do I require ATAS certificate?	Australia is one of the Countries of the World whose nationals are exempted from providing an ATAS.	✗	✓	✓	✓	✓	✓
I am from Guyana. Does that fulfil proof of English requirement for my visa process?	Guyana is one of the Countries of the World where majority of the population speaks in English or A majority English speaking country.	✗	✓	✓	✓	✓	✓
My passport is expiring, is this a problem?	Passport of the applicant should be valid at the time of application. The further criteria also depends upon the location of visa application.	✓	✓	✓	✓	✓	✓
I have a criminal record. Does that make me credible for SW visa?	A valid Criminal record certificate if specified conditions are true is one of the Eligibility requirements for a Skilled Worker Visa.	✗	✗	✗	✗	✗	✗
I have a criminal record. Will that lead to denial of my visa?	A valid Criminal record certificate if specified conditions are true is one of the Eligibility requirements for a Skilled Worker Visa.	✗	✗	✗	✓	✗	✗

References

1. Wakisaka, D., Cardwell, P.J. Exploring the trajectories of highly skilled migration law and policy in Japan and the UK. CMS 9, 43 (2021).

- <https://doi.org/10.1186/s40878-021-00251-3>
2. Manning, G. and Collinson, J., 2019. Complexity in the Immigration Rules: politics or the outcome of Judicial Review?. *Judicial Review*, 24(2), pp.85-92.
 3. Dong*, B., Yu, H. and Li, H., 2022. A Knowledge Graph Construction Approach for Legal Domain. [online] Doaj.org. Available at: <https://doaj.org/article/1171aa4f67054b39802e54b8d18d5a90> [Accessed 8 April 2022].
 4. Gov.uk. 2022. Welcome to GOV.UK. [online] Available at: <https://www.gov.uk/> [Accessed 8 April 2022].
 5. GOV.UK. 2022. Archive: Immigration Rules. [online] Available at: <https://www.gov.uk/government/collections/archive-immigration-rules> [Accessed 4 March 2022].
 6. Bouziane, A., Bouchiha, D., Doumi, N. and Malki, M., 2015. Question Answering Systems: Survey and Trends. *Procedia Computer Science*, 73, pp.366-375.
 7. Bach, N., Cham, L., Thien, T. and Phuong, T., 2017. Question Analysis for Vietnamese Legal Question Answering. In: 2017 9th International Conference on Knowledge and Systems Engineering(KSE)
 8. McElvain, G., Sanchez, G., Matthews, S., Teo, D., Pompili, F. and Custis, T., 2017. WestSearch Plus: A Non-factoid Question-Answering System for the Legal Domain. In: 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. [online] SIGIR'19 Association for Computing Machinery, pp.1361–1364. Available at: <https://doi.org/10.1145/3331184.3331397> [Accessed 8 April 2022].
 9. Junior, A., C., Orlandi, F., Graux, D., Hossari, M., O’Sullivan, D., Hartz, C., and Dirschl, C. 2020. Knowledge Graph-Based Legal Search over German Court Cases. In *The Semantic Web: ESWC 2020 Satellite Events: ESWC 2020 Satellite Events*, Heraklion, Crete, Greece, May 31 – June 4, 2020, Revised Selected Papers. Springer-Verlag, Berlin, Heidelberg, 293–297. DOI: https://doi.org/10.1007/978-3-030-62327-2_44
 10. Oxford Law Faculty. 2022. Unlocking the Potential of Artificial Intelligence for English Law. [online] Available at: <https://www.law.ox.ac.uk/unlocking-potential-artificial-intelligence-english-law> [Accessed 8 April 2022].
 11. Greenleaf, G., Mowbray, A. and Chung, P., 2017. Building Sustainable Free Legal Advisory Systems: Experiences from the History of AI ; *Law. SSRN Electronic Journal*,.
 12. Kurteva, A. (2021). Implementing Informed Consent with Knowledge Graphs. In: , et al. *The Semantic Web: ESWC 2021 Satellite Events. ESWC 2021. Lecture Notes in Computer Science()*, vol 12739. Springer, Cham. https://doi.org/10.1007/978-3-030-80418-3_28
 13. López, M.F., Gómez-Pérez, A., Sierra, J.P., & Sierra, A.P. (1999). Building a chemical ontology using methontology and the ontology design environment. *Ontologies*, 14, 37–46.
 14. Research, S., 2022. protégé. [online] Protege.stanford.edu. Available at: <https://protege.stanford.edu/> [Accessed 8 April 2022].
 15. Gate.ac.uk. 2022. GATE.ac.uk - index.html. [online] Available at: <https://gate.ac.uk/> [Accessed 8 April 2022].
 16. Wu, Z. and Palmer, M., 1994. Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics. Las Cruces, New Mexico: Association for Computational Linguistics, pp.133–138.