ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA241067

Match'In - Pilot Project of an Algorithm-Based Decision Support System for Individualized Recommendations of Municipalities for the Integration of Refugees

Christian Sauer[a], Carsten Wenzel[a], Katharina Euler[a], Sonja Reinhold[b], Frank Wuttke[c] and Björn Oitmann[c]

[a]University of Hildesheim
[b]University of Erlangen–Nuremberg
[c]eEvolution Vertrieb GmbH

Abstract. For the successful integration of refugees, it is essential to place refugees in a municipality that best matches their personal needs and future plans. It is therefore important to individualize the choice of location with the highest possible match to personal attributes and needs of a refugee. In this paper, we describe the development of the Match'In system, a recommender system that recommends municipalities for the integration of refugees, based on the knowledge representing an individual refugee and his or her needs. We describe the knowledge elicitation and formalization process we developed and applied in the Match'In project. A participatory and multi-stage process was used to develop the algorithm-based decision support system for the individualized recommendation of municipalities for refugees. We detail on the decision and rational of using case-based reasoning (CBR) as the reasoning approach within the Match'In system. We then describe the specific capabilities of the use of structural CBR in terms of transparency of the knowledge model, with an emphasis on the detection and prevention of algorithmic bias. We further describe the reasoning technique and the possibilities to automatically derive explanations from these to justify the recommendation of specific municipalities by the Match'In system.

1 Introduction

Refugees and asylum seekers within the regular asylum procedure in Germany cannot choose their exact place of arrival themselves. Instead, refugees in Germany are, in a first step, distributed to the federated states and, in a second step, to the municipalities within their territories. Both decisions are mainly based on a quota system. The individual needs and characteristics of refugees, as well as the infrastructure and resources of the local authorities, are not systematically considered when this decision is made. This regularly creates 'mismatches', causing dissatisfaction for both refugees and municipalities. The current system is often criticized, putting pressure on the institutions in charge to seek alternatives.

In recent years, algorithmic matching systems have begun to play a key role in reforming the placement of refugees in several countries, usually as part of resettlement schemes, for an overview see Ozkul [18]. The existing projects have a strong focus on the labor market. However, taking a broad understanding of integration as a basis, long-term integration depends on a broad variety of factors, as is e.g. emphasised by Ager and Strang [2]. This is the case both on the side of the refugees and of the receiving communities, as Bansak et al. [7] have shown. Therefore, the Match'In[1] project strives to develop a decision support system for matching purpose that includes many different criteria of refugees and municipalities, aiming at the best possible integration of the refugees by accommodating their needs and resources in more dimensions than only the labor market. A successful integration is not only for the benefit of the refugees but also for the receiving municipalities, saving resources and strengthening local communities.

This goal cannot be reached without the use of an algorithm, as the criteria that need to be taken into account are far too complex for the administrative staff to consider when making a placement decision. For that reason, within the Match'In project [13], scientists from both social and computer science collaborate to develop an algorithm-based matching system. The algorithm will then be used to deliver decision support by recommending municipalities for the institutions responsible for the placement of refugees in German municipalities.

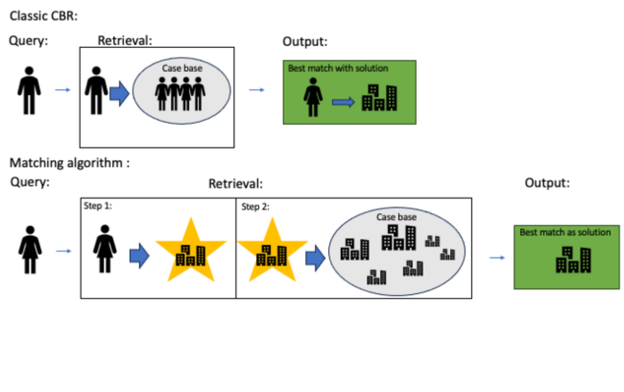

Figure 1. Matching recommendation algorithm workflow

In this paper, we first formulate four research questions on the possibilities, challenges and possible limitations of creating and using an algorithm-based decision support system for individualized

[1] https://matchin-projekt.de/en/about-the-project/

recommendations of municipalities for the integration of refugees in section 2 of the paper. To answer these four research questions, we first review relevant work on the use of knowledge-based systems and give a brief introduction into matching algorithms, the AI technique of case-based reasoning (CBR), as well as the open source CBR development tool "myCBR" in section 3. In section 4 we describe the interdisciplinary approaches we employed for the elicitation and formalization of the domain knowledge to create the knowledge model on the Match'In system operates. We then describe the Match'In algorithm itself and its use of CBR as well as the possibilities of generating explanations for the recommendation of municipalities by the Match'In algorithm in section 5. Section 6 then gives an overview of the software implementation and integration of the operational Match'In system. In section 7 we describe the currently ongoing evaluation of the Match'In algorithm and system and in section 8 we conclude on the current Match'In system and give a brief overview of the next steps we plan for the Match'In system.

2 Research questions

This paper addresses the following research questions that had to be answered as part of the development of an algorithm for the distribution of refugees to municipalities in 4 different federal states in Germany:

Research question 1. How can highly complex domain knowledge be formalized, abstracted for a matching procedure so that correlations are retained with minimal loss of knowledge?

Research question 2. How can this matching process be performed algorithmically?

Research question 3. How can the algorithm provide explanations for recommendations provided?

Research question 4. What can a bias-free knowledge model look like?

3 Relevant background

3.1 Use of knowledge-based systems for decision support

A technical definition of a knowledge-based system (KBS) is provided by Chandrasekaran, Johnson and Smith as a "KBS having explicit representations of knowledge as well as inference processes that operate on these representations to achieve a goal". Furthermore "an inference process consists of a number of inference steps, each step creating additional knowledge. [...] typically, both domain knowledge and possible inference steps have to be modeled and represented in some form."[9]. There are a number of categories into which certain types of KBS can be classified. The five main types of KBS are Expert systems, Linked systems, CBR systems, Database systems in conjunction with an intelligent user interface and intelligent tutoring systems [3]. Frequent objectives of KBS are the facilitation of tasks for non-expert users, providing decision support, the recommendation of products and workflows, strategy and planning support, design support, providing interactive tutorial systems as well as providing financial analysis[12]. On a more abstract level KBS are employed to solve problems of the following kind: Diagnosis, Selection, Prediction, Classification, Optimization, Control [17].

3.2 Matching algorithms

In the Match'In algorithm, case-based reasoning (CBR), see section 3.3, is used in such that the criteria describing a refugee's preferences, situation, and needs and the criteria that describe the facilities and characteristics of receiving communities are taken into a synthesis process that enables them to be modeled in the same way. This approach enables that a description (case) of a refugee's preferences, situation, and needs can be matched to the most similar (case) description of the facilities and characteristics of receiving communities. This approach is further supported by a number of rule-based decisions in the pre-processing of criteria describing the refugee and mapping them to criteria of the receiving communities.

Next to the matching or AI technique of CBR we evaluated further matching techniques, such as the Gale-Shapley algorithm, propensity score matching, and Mahalanobis matching. However, after evaluating the applicability of the previously named matching algorithms on the knowledge-base we are working on in the Match'In project, we concluded that CBR is the best applicable matching algorithm. The knowledge model that the Match'In algorithm employs to deduce its recommendations needs to be a transparent knowledge model, e.g., it must be fully traceable and explainable how the algorithm created a recommendation. This transparency is important to prevent hidden algorithmic bias [11], which needs to be avoided strictly. In a transparent knowledge model such bias can be identified and avoided by adjusting the knowledge model. In an opaque knowledge model, such as an artificial neural network, this kind of transparency is not possible and hence it is no suitable technique for our purpose. Further, the transparency of the knowledge model allows for explanations being automatically generated by the algorithm, based on the knowledge model, following the approach of explainable AI (XAI) [4].

3.3 Case Based Reasoning

The basic idea of case-based reasoning is to imitate a human problem-solving strategy. When we are faced with a new problem in person, we try to remember similar problems so that we can apply the solution to the problem we have already solved to the problem that is new to us.

This procedure is formalized in the so-called CBR-cycle shown in Figure 2 introduced by A. Aamodt and E. Plaza [1]. The process of the CBR-cycle provides four successive steps for the process of being applied to a new problem: 1.Retrieve, 2.Reuse, 3.Revise and 4.Retain.

Knowledge about the problem domain as well as general knowledge is stored in the knowledge base of the system. This knowledge base also contains the so-called case base. In the case base, problems are stored with their description and the corresponding solution in the form of so-called cases. Furthermore, so-called local and global similarity measures for comparing problems are stored in the knowledge base [5]. In addition, the knowledge base can store knowledge in the form of rules, e.g. completion rules and/or association rules [16],[6].

In the first step of the retrieval process, the problem description is entered as a query case in a formalized manner and, if necessary, preprocessed by the system. For example, completion rules are used to improve a query. Furthermore, within the case base, the query case is compared with the cases in the case base. Depending on the application domain, either one or more cases from the case base are retrieved based on their similarity to the query case. In the reuse step, the solution of the most similar case is considered as a possible solution for the new problem. General knowledge from the knowledge base can be used to validate this solution. In the revise step, general knowledge from the knowledge base can be used to validate this solution. Alternatively, the revise step can also be performed by a human or

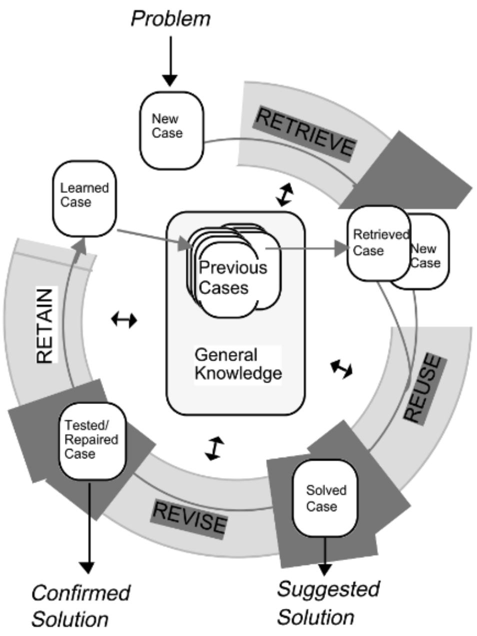

Figure 2. The CBR Cycle according to Aamodt and Plaza.

by evaluating the solution automatically. Once the solution has been validated, it is saved as a new case in the case base together with the new problem description. This step is termed the retain step.

3.4 Selection of the open source CBR tool "myCBR"

After evaluating common CBR tools, we selected the open source application myCBR [2]. The Vocabulary in myCBR 3 is stored in the model component, using various types of attribute descriptions. The Vocabulary can either be built from scratch or can rely on the import of existing data. As myCBR 3 is focused on the creation of structural CBR systems, the case representation is either flat or object oriented [6]. Furthermore, the myCBR 3 Workbench aids the knowledge engineer in the creation of the vocabulary by providing facilities to define value ranges for numeric and symbolic attributes [5]. These facilities have proven to be very helpful in the discussion with domain experts, while formalizing the vocabulary, during this research work. The vocabulary was subject to research in the implementation of this research work.

The Similarity measures are represented independently from the Vocabulary and are connected to the model itself. The data types covered by the similarity measures in myCBR 3 are simple data types like Integer, Float, Double for numeric attributes and String for textual attributes. The formats the similarity measures can be modeled in are simple and complex functions (with respective GUIs) for the numeric data types and comparative tables as well as taxonomic modeling for String (Symbol) Attributes [5, 23]. A GUI that integrates the data model is available for retrieval within the myCBR workbench. Alternatively, there is a REST interface for the integration of myCBR into web-based applications. This enables the generation of an Open API client for communication with the myCBR API running on a Server.

[2] https://mycbr-project.org/

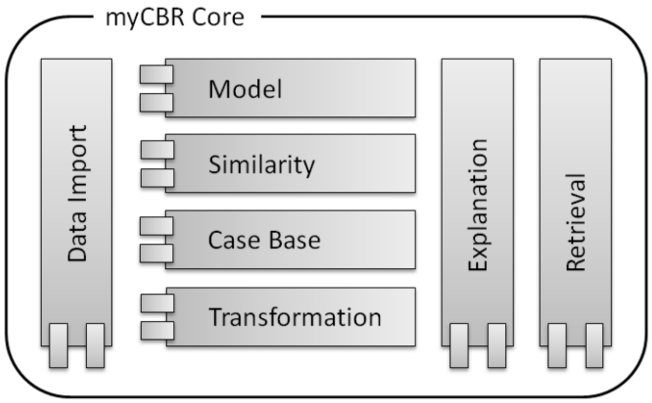

Figure 3. The myCBR Architecture [5]

4 Domain knowledge formalization

4.1 Formalizing integration factors of refugees and municipalities

In this section we review the process of bridging the gap between theory and practice by formalizing integration factors of refugees and municipalities to be used algorithmically. The process of knowledge formalization took place within the interdisciplinary team of the Match'In project in order to make the social science domain knowledge on integration and its success factors usable for an algorithm.

Step 1: Collection and structuring of relevant matching criteria First, the social sciences determined, on the basis of existing research, which factors are decisive for integration and therefore need to be taken into account when making more accurate distribution decisions with the goal of strengthening local integration. In order to supplement this data, the research gaps were filled with our own qualitative data, based on expert interviews and focus groups.

Example: "It is highly important that refugees arriving in the municipality are accommodated according to their special protection needs."

Subsequently, the relevant criteria were divided into main and subcriteria and included in a list, the scope of which was condensed in several steps. This was done in iterative steps in exchange with practitioners and experts, arriving at a first final list of matching criteria.

Example: "The relevant special protection needs apply to victims of sexual violence, victims of torture and victims of human trafficking, but also unaccompanied minors, women traveling alone, LGBTIQ people, and people with disabilities and serious or chronic illnesses".*

The main criteria chosen for the matching on the side of the refugees are personal details (e.g. gender), living (e.g. wish for decentralized housing), work (e.g. previous experiences), education/qualification (e.g. degrees), hobbies/interests (e.g. leisure activities), personal preferences (e.g. connections to linguistic communities), and special protection needs (e.g. disabilities). These are matched with the following main criteria on the municipalities' side: living (e.g. availability of accessible housing), language courses (e.g. job-related German classes), work (e.g. employment ratios), education (e.g. universities), health (e.g. hospitals), culture and leisure (e.g. sports clubs), living environment (e.g. closeness to nature), mobility (e.g. public transport), and support and advice services (e.g. migration counseling). These criteria comply with the factors identified as being decisive for a successful integration, including hard facts (like the labour market) as well as softer criteria (like hobbies), building on the knowledge that both are equally important.

Choosing these criteria, even considering the participatory and multi-stage process, brought with it an inevitable reduction of complexity. Refugees, like all human beings, have a complex set of individual skills, needs, and wishes. By breaking these down into measurable indicators, we had to find a compromise between recognizing peoples' individuality and choosing criteria that can be processed by an algorithm. In order to include as many important perspectives as possible in the process, we conducted focus groups with refugees, NGOs representing their interests, representatives of municipalities, experts for specific protection needs, and actors currently working in the area of refugee allocation within the respective authorities. This way, we could not only determine which factors to include, but also how much weight should be put on certain criteria (cf *Step 3*).

Example: "In a focus group with refugees, one participant named her teaching job as the significant factor for settling in, another referred to his football club as the most important element, as it helped him in other areas of life, e.g. acceptance into a university and finding a place to stay."

Step 2: Operationalization The social science knowledge of the consolidated list of criteria was then translated into technically usable, collectible and comparable indicators. One or more indicators were created for each criterion. The various characteristics were defined in the form of response options for the questionnaires[3] which are available online via the URL given in the footnote.

Example: "The following indicators on the side of the municipalities are queried for the prototype: Suitable accommodation for 1) women traveling alone, 2) men traveling alone, 3) LGBTIQ persons, 4) People with an increased need for psychological care."*

Step 3: Weighting of the matching criteria and criteria synthesis Drawing from data gathered in the first steps, the criteria were weighted, hence their influence in comparison with other criteria for the allocation recommendation was individually defined, again using an iterative feedback process.

Weighting The process of weighting the criteria is not only part of the algorithmic design as such but already forms one element of the project's data collection process. Hence, a significant pre-step of the weighting is already carried out by structuring the criteria into three types during data collection on the refugee side: exclusion or dropout criteria, possible exclusion criteria, and matching criteria, as shown in **figure 4.**

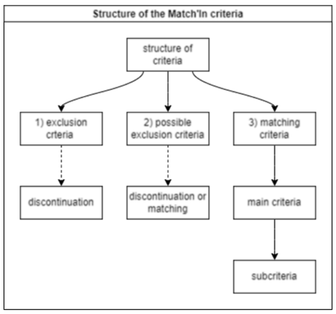

Figure 4. Structure of the Match'in criteria

1) Exclusion criteria that lead to discontinuation: Participants are first asked about the aspects that could potentially stand in the way of their participation in the matching. These exclusion criteria result in particular from legal obligations (such as the right to be allocated to close family members in a German municipality) and from the fact that those seeking protection should not suffer any disadvantages as a result of participating in the project.

2) Possible exclusion criteria: Discontinuation of participation is possible, but not absolutely necessary - the criteria only lead to discontinuation and direct recommendation to the respective municipality if the person explicitly requests this themselves.

3) Matching criteria: If participation has not been excluded, all other criteria are collected as matching criteria divided into main criteria and sub-criteria and included in the distribution decision with their respective weights. Their weighting within the algorithm and their collection, again, follows four basic rules:

A) All main criteria are weighted equally by the algorithm. Accordingly, these categories each have the same weight for the algorithm's distribution proposal.

B) The sub-criteria are weighted within the respective main category.

C) At least one sub-criterion under each main criterion must be answered as part of the data collection.

D) Special protection needs play a separate role due to important legal and ethical standards: Here, in contrast to the other main criteria, each applicable sub-criterion is weighted individually, not the entire area in a fixed sum. The individual weights of the sub-criteria under special protection needs are added together and thus result in an individual sum that is based on the specific needs of a person. This procedure is necessary in order to take into account the possible existence of mutually reinforcing, intersectional special protection needs (eg. a woman traveling alone who is also pregnant, see below in **figure 5**).

main critieria	weighting	sub-criteria (examples)	weighting
personal data	10 / 10		
living	10 / 10		
work	10 / 10		
		past professional activities / work experience	5
		professional aspirations	4
education / qualifications	10 / 10		
hobbies / interests	10 / 10		
preferences	10 / 10		
special protection needs	individual weight, eg. here: 14		
		LGBTIQ*	8
		older age	8
		pregnancy	6
		serious physical illness	9
		need for psychological treatment	9
		physical limitations	between 6 and 10
		cognitive limitations	7
		women travelling alone	8

Figure 5. Example weighting of the matching criteria

4) Individual weighting (preference-based element): As part of the data collection surveys, the participants are given the opportunity to go beyond the standard weighting by the algorithm and give double weighting to a particularly important category for them individually. This creates the possibility for participants to include their individual preferences.

Criteria synthesis

As part of the criteria synthesis, the indicators on the side of refugees were set in relation to those on the side of the municipalities. An indicator on the refugee's side can relate to one or more municipality indicators.

Example: "1) Weighting: On a scale of 1 to 10, special protection needs were weighted between 6 and 10. Inter alia, 'women traveling alone' was weighted at 8. 2) Criteria synthesis: It relates to the

[3] https://hilpub.uni-hildesheim.de/handle/ubhi/16960

following criteria on the side of the municipalities: Accessibility of women's counseling centers, women-specific projects and special requirements for the protection against violence in accommodations within the local authority a person is to be allocated to."

During the criteria synthesis we developed a set of 64 rules that are based on the necessities derived from the characteristics and needs of refugees to setup a hypothetical municipality that would match the needs of a specific refugee in an ideal way.

Example: IF: "Difficulty walking or climbing stairs" == 1 OR == 2 OR == 3 THEN: IF: "Need/desire for centralized or decentralized accommodation (expectation management, weighting)" == 1: Set: "Barrierfree_low_barrier_housing_GU" = 100

In the above example the criterion "Barrierfree_low_barrier_housing_GU", describing the availability of barrier-free or low-barrier housing in a municipality, is set to the maximum possible value to match the need of a refugee if he/she has indicated that he/she has difficulties walking or climbing stairs.

A key aspect of the use of the formalized knowledge, especially the information about individual refugees, is the provision of secure data protection as well as the use of anonymization techniques if personal data of a refugee has to be used by the Match'In system. To address these key points, we decoupled personal data of a refugee from the description of an individual refugee's needs. We further defined value ranges for a number of criteria of a refugee to prevent the re-identification of a person by the analysis of precise values, such as age or specific data on a refugee's education.

4.2 Knowledge model

One of the primary reasons for making the decision in favor of a CBR-based approach is the potential to utilize various forms of knowledge representation such as ontology, taxonomy, similarity measures for numerical values, rules, etc. [19]. As myCBR [15] can serve for both reasoning and modeling, it is also integrated into our modeling process, thus obviating the necessity for model adaptation at runtime.

The final knowledge model that we developed from the knowledge elicitation process contains 87 criteria that describe a municipality. For each of the 87 criteria, a specific similarity measure was developed. The criteria of the municipality are further grouped into the following groups: living, language courses, work, education, health, culture and leisure, living environment, mobility, and support and advice services. The grouping of the criteria of a municipality allows to apply weights to specific groups and thereby increase their influence on the recommendation of a municipality, based on the needs of an individual refugee. Furthermore, we derived a set of 64 rules, an example for such a rule is given in section 4.1, that are applied to generate an ideal municipality based on the specific needs of a refugee which are entered into the system using the developed questionnaire (see section 4.1).

In conclusion, we have shown in this section that it is possible to formalize the complex domain knowledge given in the domain of allocating refugees to best fitting municipalities, answering research question 1: How can highly complex domain knowledge be formalized, abstracted for a matching procedure so that correlations are retained with minimal loss of knowledge?

5 The 'Match'In' algorithm

The Match'In algorithm is a two-stage process based on the CBR cycle. In Chapter 3.3, the fundamental terms and functions of a CBR system were introduced. These components are represented in the Match'In algorithm as follows: a query represents the problem statement and includes the machine-readable form of the information about a refugee and form their query corpus. The general knowledge includes rules for examining and adjusting the query, and the case base consists of a set of potential target municipalities. The basic structure of CBR makes it possible to provide explanations for suggestions made by the algorithm and, if necessary, to enrich them with knowledge from the knowledge base. The two steps of the algorithm are described in the next two sub-sections. Furthermore, a third sub-section deals with how explanations can be prepared by the algorithm for the user and how these can support an end user if necessary.

5.1 Mapping refugees to a hypothetical ideal municipality

Case-based reasoning offers the use of completion rules, which are originally intended to enrich the query with knowledge from the knowledge base, e.g. to add missing but conclusive attributes from the query to the query. In the first step of the Match'In algorithm shown in Figure 1, a set of rules defined by experts (see section 4: Domain knowledge formalization) are applied to the query consisting of data from a refugee. On the one hand, these rules are used to check whether a recommendation for the distribution of the refugee can be offered by the algorithm; exclusion criteria may be present here. To clarify, this of course means that an allocation still takes place. On the other hand, the rules are used to form an "ideal" municipality for a refugee that is described as completely as possible. An ideal municipality is described as one that fulfills all the needs of a refugee. Another advantage provided by CBR as a method of argumentation is worth highlighting at this point as well: Requests for a CBR system do not have to be complete. That is, if a refugee cannot or does not want to answer all the questions on the questionnaire for the algorithm, an ideal municipality can still be generated from the request. However, this ideal municipality would only consider the knowledge provided to it. By utilizing these rules in the form of completion rules, our request is prepared to the extent that it is available for the second step of the matching algorithm.

5.2 Using CBR to identify best matches

In the second phase of the algorithm, the actual retrieval phase, the fundamental principles of retrieval from classical case-based reasoning are adopted, but adjustments are made in the representation of cases. In contrast to classical CBR, where a case consists of a description of attribute values plus a solution, in our algorithm, the case itself, which also consists of attribute values, is a possible solution. Furthermore, our algorithm allows for indirect influence of refugees on their selection of potential recommendations. During the interview phase, refugees can specify which aspects are of particular interest to them in their location choice. These attributes are assigned a higher global weight on average. This is intended to emphasize participation, but a single permitted input is avoided to prevent distorting the results until, for example, a desired outcome is achieved. Additionally, we aim to prevent the formation of bias within the system with this approach.

During the retrieval process, the ideal municipality formed in Step 1 for a refugee is compared to all cases in the case base, e.g. the available real municipalities, based on their attribute values using similarity measures stored in the CBR system. The algorithm outputs the X municipalities most similar to the 'ideal' municipality as suggestions for potential assignments of the asylum seeker. Where X is a number of 1 to all available real municipalities, sorted in a decreasing order by their degree of similarity to the generated hypothetical ideal municipality.

In the Reuse step of the CBR cycle, the results of retrievals are adopted as a solution for an assignment option for the refugee. The Revise step is carried out by the case worker of the responsible authority, who uses the suggestions of the Match'In algorithm, among other things, as a basis for a decision on the assignment. It has been decided against direct feedback or returning the decision on the assignment to the system, which would take place in the Reuse step. New cases or new knowledge in the knowledge base must be explicitly entered into the system by an admin user. So in summary, in this section we have answered the research question 2: "How can this matching process be performed algorithmically?" by presenting the workflow and techniques the Match'In algorithm employs.

5.3 Generating explanations for decision support

Following Roger Schank [22], explanations are the most common method used by humans to support their decisions. Therefore, an explanation aims to explain a solution to a problem and the reasoning that led to this solution. Furthermore, an explanation must be able to point out how a system works and how it is operated [20]. Next to these purposes of an explanation, a number of more subtle but nevertheless not less important goals of explanations exist. A user's trust in the system and its results can also be increased by the provision of explanation on the results themselves as well as on the reasoning process of the system that led to these results [14]. Furthermore, explanations can be used by a knowledge-based system to justify its solutions and thereby becoming accountable for its solutions [10]. Explainable AI, in short XAI, meaning the ability of AI systems in general to explain their reasoning and so justify their decisions or in our case recommendations, is currently a core topic of research in AI systems [4].

The research question 3 of this paper is "How can the algorithm provide explanations for suggestions provided?". To answer this research question, we will now describe the approach of generating explanations in the Match'In system. For the Match'In system the explanatory knowledge, the knowledge needed to generate explanations [8], consists of the CBR knowledge model's knowledge containers, namely the vocabulary, the cases, the 87 similarity measures and the adaptation knowledge of the CBR knowledge model [21] and the set of 64 rules derived during the knowledge formalization process of the domain knowledge, see section 4: Domain knowledge formalization. The fact that structural CBR uses a completely transparent knowledge model as well as a completely transparent reasoning technique, which means the knowledge model as well as the reasoning techniques can be retraced by a human user, was a major factor for our decision to use structural CBR for the Match'In system. The complete transparency of the knowledge model and reasoning techniques in the Match'In system also allows for easy detection of possible algorithmic bias [11] and a subsequent correction of either the knowledge model or adaptation of the reasoning techniques. With regard to research question 4: "What can a bias-free knowledge model look like?" we can state that it is imperative that the knowledge model used must be 100% transparent to allow for the detection and, if necessary, correction of algorithmic bias. For the next iteration of the Match'In system it is planned to generate explanations for the purpose of justifying its recommendations. Therefore it will use the 87 similarity measures and an analysis of the firing of the 64 rules from the rule set. Analyzing these can provide a fully transparent retrace of the reasoning that led to the system's recommendation.

It is further planned to derive short textual explanations from the firing of the rules as well as to add short textual explanations derived from the application of the similarity measures of the CBR knowledge model. Additionally, though not aimed on the justification of the recommendations of the system, the Match'In system currently uses "canned explanations", pre-defined texts that were elicited and prefabricated during the knowledge formalization process, see section 4, to support the user of the system via its GUI.

6 Technical implementation

The development of the Match'In system leverages a groundbreaking integration of two distinct technological systems: a database-supported web application for data capture, and a case-based reasoning (CBR) matching system implemented using myCBR. The web application is crafted with ASP.NET Blazor Server and a LowCode technology, while the matching system is developed in Java, connected via a RESTful interface. Despite the divergence in their technological stacks, a specially designed C# library bridges these systems efficiently, utilizing the myCBR REST API to provide functionalities tailored for the project's specific CBR model.

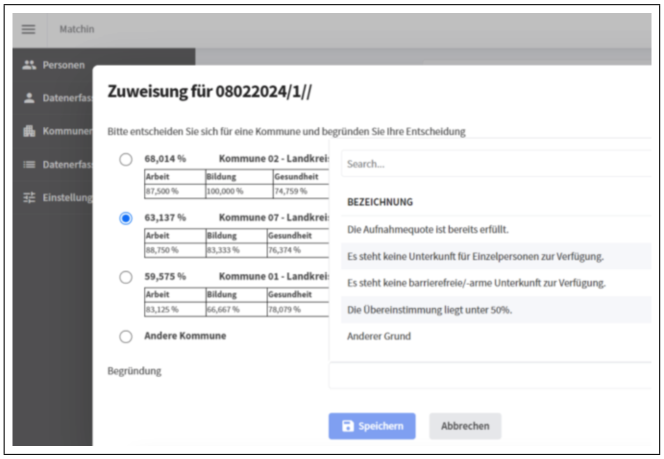

Figure 6. Example view of the software (proposal selection)

Within the web application framework, a myCBR Service has been integrated to streamline the access and updating processes for the database, enabling the system to refresh or establish new municipality information based on the refugees' profiles. This service facilitates the identification of ideally matched municipalities for individuals seeking refuge, thereby optimizing the allocation process. The implementation encompasses a variety of communication and data classes, ensuring seamless integration and manipulation of complex attributes and values conversion between the disparate systems.

The matching process initiates once an individual seeking refuge completes a questionnaire, triggering the web application to retrieve pertinent information about municipalities from the database and synchronize it with the myCBR model. The system evaluates whether each municipality is already represented within the model or needs to be added. Subsequently, it constructs an "ideal" municipality based on the individual's responses, a task orchestrated by a sophisticated rule class known as the 'PersonenAnfrageBuilder'. This class is responsible for translating the detailed responses from the questionnaire into a comprehensive set of attributes that define the ideal municipality. Utilizing a series of predefined rules, derived from extensive research and collaboration with experts, the 'PersonenAnfrageBuilder' adjusts and refines these attributes to accurately reflect the needs and preferences of the refugee. This meticulous process ensures that the ideal municipality constructed is a close match to the individual's requirements, taking into account factors such as availability of services, community support, and other critical amenities that can significantly impact the well-being of the refugee.

The system then queries for a list of municipalities that best match the idealized profile, storing the outcomes in the database. Based on these results, municipalities are ranked in descending order of their overall match quality, thereby enabling an effective and equitable distribution of refugees to the most suitable locations.

This technical solution exemplifies the potent synergy between web technologies and artificial intelligence in addressing and resolving the challenges associated with refugee distribution. The architecture's flexibility and the system's efficient data processing capabilities are pivotal in enhancing decision-making processes and promoting fair distribution practices. By harnessing the power of case-based reasoning and sophisticated web application frameworks, the project showcases a novel approach to leveraging technology for the social good: paving the way for improved integration and support mechanisms for refugees worldwide.

7 Evaluation

The first iteration of the Match'In system was implemented and is currently in evaluation in a pilot phase in participating authorities in four federal states of Germany, namely Hesse, North Rhine-Westphalia, Rhineland-Palatinate and Lower Saxony. The current case-base of the Match'In system contains 20 pilot communities distributed over the named four federal states. Prior to this, an expert evaluation of the results was carried out using test data. The system was able to show that suitable municipalities could be recommended as possible proposals for the refugees to be assigned to. We are currently in the pilot phase of our project, primarily focusing on data gathering. This phase is crucial for viable suggestions of further refinement and development of the Match'In system. One of our key activities during this pilot phase is the ongoing evaluation of the recommendations generated by our knowledge model and algorithm. We are meticulously assessing their accuracy and relevance to ensure the effectiveness of our system. Additionally, we are actively soliciting feedback on the usability of the Match'In software from local authorities across the 4 federal states that participate in the pilot phase. This feedback is invaluable as it helps us refine and improve the user experience, ensuring that our software meets the diverse needs of our stakeholders. To fascinate these activities, we are in direct exchange with the local authorities in the four federal states and further use additional questionnaires as well as focus group meetings and interviews with all involved parties. A further focus of our evaluation also lies on the assessment and improvement of the acceptance of the questionnaire used by refugees to characterize their needs and circumstances. The evaluation of the questionnaire is primarily carried out by interviewing the participating refugees and the counselors in the reception facilities who enter the participating refugees' data into the software. We just started the pilot phase of the Match'In system in February 2024 and therefore are currently in the data and feedback gathering stage of the evaluation, using the techniques and involving the stakeholders described above.

8 Conclusion and Outlook

In this paper we first formulated four research questions on the possibilities, challenges and possible limitations of creating and using an algorithm-based decision support system for individualized recommendations of municipalities for the integration of individual refugees. To answer these four research questions, we reviewed relevant work on the use of knowledge-based systems and gave a brief introduction into matching algorithms, the AI technique of CBR and the CBR development tool "myCBR". We further described the interdisciplinary approaches we employed for the elicitation and formalization of the domain knowledge to create the knowledge model on which the Match'In system operates. We also described the Match'In algorithm itself in detail. Furthermore, we gave an overview of the software implementation and integration of the operational Match'In system. We were able to answer research question 1 in section 4 of this paper, research question 2 in section 5.2 of this paper, research questions 3 and 4 in section 5.3 of this paper. As we progress beyond the pilot phase of the Match'In system, our next steps involve a deeper evaluation of the system's performance and functionality. This includes refining our knowledge model to enhance its accuracy. Moreover, we plan to expand the usage of our system beyond the pilot municipalities. This wider implementation will allow us to gather more diverse data and insights, enriching the effectiveness of our solutions. Furthermore, we aim to enhance the explanatory capabilities of our system. By refining the generation of explanations, we can provide users with clearer insights into the rationale behind recommendations, fostering better understanding and trust in the system. In summary, our focus remains on continuous improvement, with a commitment to refining our knowledge model and algorithm, expanding usage of the Match'In system, and enhancing explanation generation for a better and user-friendly experience.

Acknowledgements

The Match'In project (https://matchin-projekt.de/en/about-the-project/) is supported by the Mercator Foundation Germany. The authors would also like to thank the involved actors of the authorities of the federal states of Hesse, North Rhine-Westphalia, Rhineland-Palatinate and Lower Saxony for their participation in the Match'In project.

References

[1] A. Aamodt and E. Plaza. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Commun.*, 7(1): 39–59, mar 1994. ISSN 0921-7126.

[2] A. Ager and A. Strang. Understanding integration: A conceptual framework. *Journal of Refugee Studies*, 21, 04 2008. doi: 10.1093/jrs/fen016.

[3] R. Akerkar and P. Sajja. *Knowledge-based systems*. Jones & Bartlett Publishers, 2009.

[4] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58: 82–115, 2020.

[5] K. Bach. *Knowledge Acquisition for Case-Based Reasoning Systems, PhD thesis*. Universität Hildesheim, 2013.

[6] K. Bach and K.-D. Althoff. Developing case-based reasoning applications using mycbr 3. In *Cased-Based Reasoning Applications and Research*, volume 7466, pages 17–31, 09 2012. ISBN 978-3-642-32985-2. doi: 10.1007/978-3-642-32986-9_4.

[7] K. Bansak, J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, and J. Weinstein. Improving refugee integration through data-driven algorithmic assignment. *Science*, 359:325–329, 01 2018. doi: 10.1126/science.aao4408.

[8] B. Chandrasekaran and W. Swartout. Explanations in knowledge systems: the role of explicit representation of design knowledge. *IEEE expert*, 6(3):47–49, 1991.
[9] B. Chandrasekaran, T. R. Johnson, and J. W. Smith. Task-structure analysis for knowledge modeling. *Communications of the ACM*, 35(9):124–137, 1992.
[10] D. Doyle, A. Tsymbal, and P. Cunningham. A review of explanation and explanation in case-based reasoning, 2003.
[11] S. Fazelpour and D. Danks. Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8):e12760, 2021.
[12] G. Jakus, V. Milutinović, S. Omerović, S. Tomažič, G. Jakus, V. Milutinović, S. Omerović, and S. Tomažič. *Knowledge representation*. Springer, 2013.
[13] Match'In project, 2021-2024. URL https://matchin-projekt.de/en/.
[14] J. D. Moore and W. R. Swartout. *Explanation in expert systems: A survey*. University of Southern California, Information Sciences Institute Marina del Rey, 1988.
[15] myCBR, 2016-2024. URL https://github.com/ntnu-ai-lab/mycbr-rest.git.
[16] G. Müller and R. Bergmann. Case completion of workflows for process-oriented case-based reasoning. In *Case-Based Reasoning Research and Development*, 10 2016. ISBN 978-3-319-47095-5. doi: 10.1007/978-3-319-47096-2_20.
[17] M. Negnevitsky. *Artificial intelligence: a guide to intelligent systems*. Pearson education, 2005.
[18] D. Ozkul. Automating immigration and asylum: the uses of new technologies in migration and asylum governance in europe, 2023.
[19] M. Reichle, K. Bach, and K.-D. Althoff. The seasalt architecture and its realization within the docquery project. In *KI 2009: Advances in Artificial Intelligence*, pages 556–563, 09 2009. ISBN 978-3-642-04616-2. doi: 10.1007/978-3-642-04617-9_70.
[20] T. R. Roth-Berghofer. Explanations and case-based reasoning: Foundational issues. In *European Conference on Case-Based Reasoning*, pages 389–403. Springer, 2004.
[21] T. R. Roth-Berghofer and J. Cassens. Mapping goals and kinds of explanations to the knowledge containers of case-based reasoning systems. In *International Conference on Case-Based Reasoning*, pages 451–464. Springer, 2005.
[22] R. Schank. *Explanation patterns: Understanding mechanically and creatively*. Psychology Press, 2013.
[23] A. Stahl and T. R. Roth-Berghofer. Rapid prototyping of cbr applications with the open source tool mycbr. In *European conference on case-based reasoning*, pages 615–629. Springer, 2008.