Maintenance of Profile Matchings in Knowledge Bases*

Jorge Martinez Gil[1], Lorena Paoletti[1], Gábor Rácz[2], Attila Sali[2], Klaus-Dieter Schewe[1]

[1] Software Competence Center Hagenberg, Softwarepark 21, 4232 Hagenberg, Austria
jorge.martinez-gil|lorena.paoletti|kd.schewe@scch.at

[2] Alfréd Rényi Institute of Mathematics, P.O.B.127, 1364 Budapest, Hungary
gabee33@gmail.com, sali@renyi.hu

Abstract. A profile describes a set of properties, e.g. a set of skills a person may have or a set of skills required for a particular job. Profile matching aims to determine how well a given profile fits to a requested profile. The research taken in this paper uses a knowledge base grounded in description logic to represent the knowledge about profiles. Thus, profiles can be defined by filters in the underlying lattice of the concepts in the TBox of the knowledge base. Matching can be realised by assigning values in [0,1] to pairs of such filters: the higher the matching value the better is the fit. Conversely, given a set of filters together with matching values determined by some human expert, the question is, whether a matching measure can be determined such that the computed matching values preserve the rankings given by the expert. In the paper plausibility constraints for the values given by an expert are formulated. If these plausibility constraints are satisfied, the problem of determining an appropriate matching measure can be solved in an order-preserving way.

1 Introduction

A profile describes a set of properties, and profile matching is concerned with the problem to determine how well a given profile fits to a requested one. Profile matching appears in many application areas such as matching applicants for jobs to job requirements, matching system configurations to requirements specifications, etc.

* The research reported in this paper was supported by the Austrian Forschungsförderungsgesellschaft (FFG) for the Bridge project "Accurate and Efficient Profile Matching in Knowledge Bases" (ACEPROM) under contract **[FFG: 841284]**. The research reported in this paper has further been supported by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry of Science, Research and Economy, and the Province of Upper Austria in the frame of the COMET center SCCH. **[FFG: 844597]**

Taking the profile just as a set of unrelated items is usually not appropriate for the problem, even though many distance measures between sets such as Jaccard or Sørensen-Dice [10] have proven to be useful in ecological applications. The reason is that many dependencies between the properties have to be taken into account. Therefore, in the human resources application area many taxonomies for skills, competences and education such as DISCO [2], ISCED [8] and ISCO [9] have been set up. On the grounds of these application-oriented dictionaries for profile matching a lattice structure for the individual properties can be assumed. This has been exploited by Popov and Jebelean in [14] defining a different asymmetric matching measure on the basis of filters in such lattices.

However, it can well be argued that the hierarchical dependencies in lattices are still insufficient for capturing the exact meaning of the properties in a profile. For instance, it is not common to request just "programming in Java" as a required skill, but it is more likely that further attributes are given such as years of experience associated with the skill, level of complexity of problems addressed with the skill, etc. Therefore, it appears favourable to not only assume a lattice structure, but to exploit sophisticated knowledge representation features for semantic matching problems as advocated by Falk, Mochol and others [3, 12]. In our research we adopt this basic assumption how to represent knowledge about properties. That is, we exploit description logics [1] as the basis for knowledge representation using a rather expressive language similar to SROIQ($\mathcal{D}$). With this the automatic classification of properties such as skills and competences can be supported, which is necessary in view of the frequent changes to such a knowledge base, which makes manual management of dependencies almost impossible.

This raises first the question how matching can be generalised from filters in lattices to knowledge bases. Using just the lattice defined by the named concepts can be used as a starting point, but it would ignore the fine-tuning of the knowledge that is obtained through the roles. Therefore, we exploit "blowing-up" roles, which means to enrich the concept lattice by inverse images defined by the roles [13]. In Section 2 we give a brief account on our general approach to profile matching in knowledge bases, formally defining a knowledge representation language and matching measures based on filters.

The second question, which is the core problem handled in this paper concerns the relationship of rankings obtained through the matching measures and the judgements of human experts. An initial idea based on formal concept analysis [6, 5, 4] was already presented in [11] aiming to enrich the knowledge base by additional concepts that would justify the judgement of the human expert. Here we investigate the learning of the matching measure. Start from the set of filters together with matching values or simply rankings determined by some human expert. Some plausibility constraints should nonetheless be satisfied to exclude unjustified bias that is grounded in the valuation of facts not represented in a knowledge base. We then use formal concept analysis on a family of binary relations on the set of filters determined by the expert rankings. We show that all these rankings permit to determine a suitable matching measure. This key contribution will be presented in Section 3. We conclude with a brief summary.

2 Profile Matching in Knowledge Bases

In this section we present the formal definitions underlying our approach to profile matching in knowledge bases. We will start with the general approach to knowledge representation, proceed with the representation of profiles, and discuss filter-based matching.

2.1 Knowledge Representation

For the representation of knowledge we adopt the fundamental distinction between *terminological* and *assertional* knowledge that has been used in description logics since decades. For the former one we define a language, which defines the TBox of a knowledge base, while the instances define a corresponding ABox.

A TBox consists of concepts and roles. In addition, we will permit the denotation of individuals as supported by SROIQ($\mathcal{D}$) [1] and OWL2 [7]. For this assume that C_0, I_0 and R_0 represent not further specified sets of basic concepts, individuals and roles, respectively. Then *concepts* C and *roles* R are defined by the following grammar:

$$\begin{array}{rcl} R & = & R_0 \mid R_0^- \mid R_1 \circ R_2 \\ A & = & C_0 \mid \top \mid \geq m.R \text{ (with } m > 0) \mid \{I_0\} \\ C & = & A \mid \neg C \mid C_1 \sqcap C_2 \mid C_1 \sqcup C_2 \mid \exists R.C \mid \forall R.C \end{array}$$

Definition 1. A *TBox* is a finite set $\mathcal{T}$ of assertions of the form $C_1 \sqsubseteq C_2$ with concepts C_1 and C_2 as defined by the grammar above.

Each assertion $C_1 \sqsubseteq C_2$ in a TBox $\mathcal{T}$ is called a *subsumption axiom*. Note that Definition 1 only permits subsumption between concepts, not between roles, though it is possible to define more complex terminologies that also permit role subsumption. As usual, we use several shortcuts:

- $C_1 \equiv C_2$ can be used instead of $C_1 \sqsubseteq C_2 \sqsubseteq C_1$.
- $\bot$ is a shortcut for $\neg\top$.
- $\{a_1, \dots, a_n\}$ is a shortcut for $\{a_1\} \sqcup \dots \sqcup \{a_n\}$.
- $\leq m.R$ is a shortcut for $\neg \geq m+1.R$, and $= m.R$ is a shortcut for $\geq m.R \sqcap \leq m.R$.

The semantics of a TBox is defined by structures.

Definition 2. A *structure* $\mathcal{S}$ for a TBox $\mathcal{T}$ consists of a non-empty set $\mathcal{O}$ together with subsets $\mathcal{S}(C_0) \subseteq \mathcal{O}$ and $\mathcal{S}(R_0) \subseteq \mathcal{O} \times \mathcal{O}$ for all basic concepts R_0 and basic roles R_0, respectively, and individuals $\bar{a} \in \mathcal{O}$ for all $a \in I_0$. $\mathcal{O}$ is called the base set of the structure.

We first extend the interpretation of basic concepts and roles and to all concepts and roles as defined by the grammar above, i.e. for each concept C we define a subset $\mathcal{S}(C) \subseteq \mathcal{O}$, and for each role R we define a subset $\mathcal{S}(R) \subseteq \mathcal{O} \times \mathcal{O}$ as follows:

$$\begin{aligned}
\mathcal{S}(R_0^-) &= \{(y,x) \mid (x,y) \in \mathcal{S}(R_0)\} \\
\mathcal{S}(R_1 \circ R_2) &= \{(x,z) \mid \exists y.(x,y) \in \mathcal{S}(R_1) \wedge (y,z) \in \mathcal{S}(R_2) \\
\mathcal{S}(\top) &= \mathcal{O} \\
\mathcal{S}(\{a\}) &= \{\bar{a}\} \\
\mathcal{S}(\geq m.R) &= \{x \in \mathcal{O} \mid \#\{y \mid (x,y) \in \mathcal{S}(R)\} \geq m\} \\
\mathcal{S}(\neg C) &= \mathcal{O} - \mathcal{S}(C) \\
\mathcal{S}(C_1 \sqcap C_2) &= \mathcal{S}(C_1) \cap \mathcal{S}(C_2) \\
\mathcal{S}(C_1 \sqcup C_2) &= \mathcal{S}(C_1) \cup \mathcal{S}(C_2) \\
\mathcal{S}(\exists R.C) &= \{x \in \mathcal{O} \mid (x,y) \in \mathcal{S}(R) \text{ for some } y \in \mathcal{S}(C)\} \\
\mathcal{S}(\forall R.C) &= \{x \in \mathcal{O} \mid (x,y) \in \mathcal{S}(R) \Rightarrow y \in \mathcal{S}(C) \text{ for all } y\}
\end{aligned}$$

Definition 3. An *ABox* for a TBox $\mathcal{T}$ is a finite structure $\mathcal{S}$, such that $\mathcal{S}(C_1) \subseteq \mathcal{S}(C_2)$ holds for all assertions $C_1 \sqsubseteq C_2$ in $\mathcal{T}$.

For the following we always consider a concept C in a TBox as representation of abstract properties, e.g. “knowledge of Java”, and individuals in the ABox as concrete properties such as the “Java knowledge of Lara”. Therefore, given an ABox a *profile* is simply a subset of the base set $\mathcal{O}$.

2.2 Profiles as Filters

Obviously, the concepts in a TBox define a lattice with $\sqcap$ and $\sqcup$ as operators for meet and join, and $\sqsubseteq$ for the partial order. So let us abstract for a moment from the specific definition of the knowledge base by TBox and ABox and assume to be given a lattice $(\mathcal{L}, \leq)$.

Definition 4. A *filter* in a lattice $(\mathcal{L}, \leq)$ is a non-empty subset $\mathcal{F} \subseteq \mathcal{L}$, such that for all C, C' with $C \leq C'$ whenever $C \in \mathcal{F}$ holds, then also $C' \in \mathcal{F}$ holds.

If $P \subseteq \mathcal{O}$ is a profile, then P defines in a natural way a filter $\mathcal{F}$ of the lattice $\mathcal{L}$ of concepts:

$$\mathcal{F} = \{C \in \mathcal{L} \mid \exists p \in P.\, p \in \mathcal{S}(C)\}$$

Therefore, for determining matching relations we can concentrate on filters $\mathcal{F}$ in a lattice.

2.3 Filter-Based Matching

In the following let $(\mathcal{L}, \leq)$ be a lattice, and let $\mathbb{F} \subseteq \mathcal{P}(\mathcal{L})$ denote the set of filters in this lattice.

Definition 5. A *relative weight measure* on $\mathcal{L}$ is a function $m : \mathcal{P}(\mathcal{L}) \rightarrow [0, 1]$ satisfying

(1) $m(\mathcal{L}) = 1$, and
(2) $m(\bigcup_{i \in I} A_i) = \sum_{i \in I} m(A_i)$ for pairwise disjoint A_i $(i \in I)$.

A *matching measure* is a function $\mu : \mathbb{F} \times \mathbb{F} \rightarrow [0, 1]$ such that $\mu(\mathcal{F}_1, \mathcal{F}_2) = m(\mathcal{F}_1 \cap \mathcal{F}_2)/m(\mathcal{F}_2)$ holds for some relative weight measure m on $\mathcal{L}$.

Note that the definition implies that also $m(\mathcal{L} - A) = 1 - m(A)$ must hold for a relative weight measure on $\mathcal{L}$ and any $A \subseteq \mathcal{L}$.

Example 1. The matching measure μ_{pj} defined in [14] uses simply cardinalities:

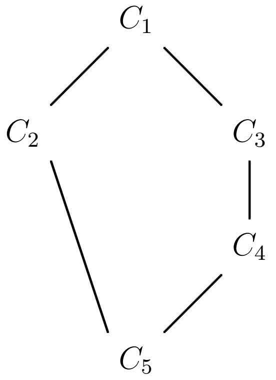
$$\mu_{pj}(\mathcal{F}_1, \mathcal{F}_2) = \#(\mathcal{F}_1 \cap \mathcal{F}_2)/\#\mathcal{F}_2$$

Thus, it is defined by the relative weight measure m on $\mathcal{L}$ with $m(A) = \#A/\#\mathcal{L}$.

It is easy to see that every matching measure μ is defined by weights $w(C) = m(\{C\}) \in [0,1]$ for the elements $C \in \mathcal{L}$. With this we immediaely obtain $m(\mathcal{F}) = \sum_{C \in \mathcal{F}} w(C)$ and thus

$$\mu(\mathcal{F}_1, \mathcal{F}_2) = \sum_{C \in \mathcal{F}_1 \cap \mathcal{F}_2} w(C) \cdot \left(\sum_{C \in \mathcal{F}_2} w(C) \right)^{-1} .$$

Example 2. Taking a simple lattice $\mathcal{L}$ with only five elements: $\mathcal{L} = \{C_1, C_2, C_3, C_4, C_5\}$. The lattice structure is shown in Figure 1.



Filters $\mathbb{F}$:

$\langle C_1 \rangle = \{C_1\}$
$\langle C_2 \rangle = \{C_1, C_2\}$
$\langle C_3 \rangle = \{C_1, C_3\}$
$\langle C_2, C_3 \rangle = \{C_1, C_2, C_3\}$
$\langle C_4 \rangle = \{C_1, C_3, C_4\}$
$\langle C_2, C_4 \rangle = \{C_1, C_2, C_3, C_4\}$
$\langle C_5 \rangle = \{C_1, C_2, C_3, C_4, C_5\}$

Fig. 1. A simple lattice and its filters

Then we obtain seven filters for this lattice, each generated by one or two elements of the lattice. These filters are also shown in Figure 1.

If we now define weights $w(C_1) = \frac{1}{10}$, $w(C_2) = \frac{3}{10}$, $w(C_3) = \frac{1}{5}$, $w(C_4) = \frac{3}{10}$, $w(C_5) = \frac{1}{10}$, then we obtain the matching measure values $\mu(\mathcal{F}, \mathcal{G}$ shown in Table 1. In the table the row label is $\mathcal{F}$ and the column label is $\mathcal{G}$.

Note that in general matching measures are not symmetric. If $\mu(\mathcal{F}_1, \mathcal{F}_2)$ expresses how well a given profile $\mathcal{F}_1$ matches a required profile $\mathcal{F}_2$, then $\mu(\mathcal{F}_2, \mathcal{F}_1)$ measures what is "too much" in the given profile $\mathcal{F}_1$ that is not required in profile $\mathcal{F}_2$ [13].

Example 3. Take the filter $\mathcal{L}$ from Example 2 shown in Figure 1. Let $\mathcal{G} = \langle C_2, C_3 \rangle$ be the filter that represents the requirements. If we take filters $\mathcal{F}_1 = \langle C_3 \rangle$ and $\mathcal{F}_2 = \langle C_4 \rangle$ representing given assets, then we have

$$\mu(\mathcal{F}_1, \mathcal{G}) = \mu(\mathcal{F}_2, \mathcal{G}) = \frac{1}{2} ,$$

	$\langle C_1\rangle$	$\langle C_2\rangle$	$\langle C_3\rangle$	$\langle C_2, C_3\rangle$	$\langle C_4\rangle$	$\langle C_2, C_4\rangle$	$\langle C_5\rangle$
$\langle C_1\rangle$	1	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{10}$
$\langle C_2\rangle$	1	1	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{4}{9}$	$\frac{2}{5}$
$\langle C_3\rangle$	1	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{3}{10}$
$\langle C_2, C_3\rangle$	1	1	1	1	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{5}$
$\langle C_4\rangle$	1	$\frac{1}{4}$	1	$\frac{1}{2}$	1	$\frac{2}{3}$	$\frac{2}{5}$
$\langle C_2, C_4\rangle$	1	1	1	1	1	1	$\frac{9}{10}$
$\langle C_5\rangle$	1	1	1	1	1	1	1

Table 1. A matching measure μ on the lattice $\mathcal{L}$

i.e., both given filters match the requirements equally well. However, if we also consider

$$\mu(\mathcal{G}, \mathcal{F}_1) = 1 \quad \text{and} \quad \mu(\mathcal{G}, \mathcal{F}_2) = \frac{1}{2} \ ,$$

then we see that $\mathcal{F}_1$ matches "better", as $\mathcal{F}_2$ contains C_4, which is not required at all.

2.4 Role-Enriched Filter-Based Matching

The matching measures introduced so far are based solely on filters in a lattice, but a TBox is more than its concept lattice. In order to fully exploit the knowledge represented in a TBox we integrate the TBox by blow-up operators. Formally, if C is a concept, for which $C \sqsubseteq \exists R.C'$ holds, then for any subconcept $C'' \sqsubseteq C'$ we can define the subconcept $\mathrm{bl}_{R,C''}(C) = C \sqcap \exists R.C''$ of C, which is called the result of the *blow-up* of R with respect to C'' on the concept C.

In particular, this becomes relevant, if C'' is defined by individuals, say $C'' = \{a_1, \dots, a_n\}$. Any subsumption between subconcepts of C' naturally induces subsumption on these blown-up subconcepts of C, i.e. we have

$$C_1 \sqsubseteq C_2 \ \Rightarrow \ \mathrm{bl}_{R,C_1}(C) \sqsubseteq \mathrm{bl}_{R,C_2}(C) \ .$$

By means of the blow-up operators we bring the information carried by the roles into additional concepts, to which the matching measures as discussed before can be applied.

At this point a few pragmatic remarks concerning the matching process in practice are due. In particular, in the model application area dealing with matching between candidate profiles (represented by filters $\mathcal{F}$)

and requested job profiles (represented by filters $\mathcal{G}$) first very few roles are used. That is, the matching will be based first only on lattices of named concepts plus their intersections and unions. Such lattices arise from the taxonomies given by DISCO [2], ISCED [8], ISCO [9] and others by adding the missing joins and meets.

In a second step a more fine-tuned matching may be required such as considering years of experience with a particular skill, level of application knowledge, etc. For this a slightly enlarged knowledge base containing roles may be exploited. Matching on top of such a knowledge would exploit the concepts resulting from the blow-up operators, but most likely the concept C' defining the co-domain of the role R to be blown-up on concept C, i.e. $C \sqsubseteq \exists R.C'$, will be defined by a finite set of individuals, say $C' = \{a_1, \dots, a_n\}$. Consequently, the enrichment of the knowledge base will mainly contain such blown-up subconcepts of C.

In a third step the "overqualification" may be considered as emphasised in Example 3. That is, profiles that equally (or almost equally) well match the requirements, may be subjected to a second ranking with respect to the inverted matching. However, it might as well be the case that concepts in the given profiles that do not appear in the requirements may as well be considered as secondary assets instead of being considered not to be needed. That is, an additional secondary matching may be based on extensions to the requested profile $\mathcal{G}$. Such a process can be repeated with tertiary and in general nth-ary matchings as long as needed.

Finally, a certain personal bias can never be fully excluded. This can be captured by additional concepts – let us call them *biases* $B_1, \dots, B_k$ – that are added to the required profile. Bias concepts may also be added to some of the given profiles to ensure that the resulting matching coincides with expectations.

3 Matching Analysis

Let $\mathcal{L}$ be a lattice with profiles defined by filters. Let $\mathbb{F}$ denote the set of all filters. Note that each filter $\mathcal{F} \in \mathbb{F}$ is uniquely determined by its minimal elements, so we can write $\mathcal{F} = \langle C_1, \dots, C_k \rangle$. The matching knowledge of a human expert can be represented be a partial mapping $h : \mathbb{F} \times \mathbb{F} \to [0, 1]$. Though human experts will hardly ever provide complete information, we will assume in the sequel that h is total.

The general question is whether there exists a matching measure μ on $\mathbb{F}$ as defined before such that $\mu(\mathcal{F}, \mathcal{G}) = h(\mathcal{F}, \mathcal{G})$ holds for all pairs of filters. As the matching values as such are merely used to determine rankings whereas their concrete value is of minor importance, this problem can be weakened to find a *ranking-preserving* matching measure μ on $\mathbb{F}$, i.e. the matching measure should imply the same rankings.

Definition 6. A matching measure μ on $\mathbb{F}$ is called *ranking-preserving* with respect to $h : \mathbb{F} \times \mathbb{F} \to [0, 1]$ iff

(1) for all filters $\mu(\mathcal{F}_1, \mathcal{G}) \geq \mu(\mathcal{F}_2, \mathcal{G})$ holds, whenever $h(\mathcal{F}_1, \mathcal{G}) \geq h(\mathcal{F}_2, \mathcal{G})$ holds, and
(2) for all filters $\mu(\mathcal{F}, \mathcal{G}_1) \geq \mu(\mathcal{F}, \mathcal{G}_2)$ holds, whenever $h(\mathcal{F}, \mathcal{G}_1) \geq h(\mathcal{F}, \mathcal{G}_2)$ holds.

3.1 Plausibility Constraints

We are looking for plausibility constaints for the mapping h that should be satisfied in the absence of bias, i.e. the assessment of the human expert is not grounded in hidden concepts. If such plausibility conditions are satisfied we explore the existence of a ranking-preserving matching measure μ. First we show the following simple lemma.

Lemma 1. *Let μ be a matching measure on $\mathbb{F}$. Then for all filters $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2, \mathcal{G} \in \mathbb{F}$ the following conditions hold:*

(1) $\mu(\mathcal{F}, \mathcal{G}) = 1$ for $\mathcal{G} \subseteq \mathcal{F}$.
(2) $\mu(\mathcal{F}, \mathcal{G}) \leq \mu(\mathcal{F}, \mathcal{G} - \{C\})$ holds for $C \notin \mathcal{F}$.
(3) $\mu(\mathcal{F}, \mathcal{G}) \leq \mu(\mathcal{F} \cup \{C\}, \mathcal{G} \cup \{C\})$.
(4) If $\mathcal{F} \cap \mathcal{F}_1 \cap \mathcal{G} = \mathcal{F} \cap \mathcal{F}_2 \cap \mathcal{G}$ holds, then

$$\mu(\mathcal{F}_1, \mathcal{G}) > \mu(\mathcal{F}_2, \mathcal{G}) \Leftrightarrow \mu(\mathcal{F}, \mathcal{F}_1 \cap \mathcal{G}) < \mu(\mathcal{F}, \mathcal{F}_2 \cap \mathcal{G}).$$

Proof. Property (1) is obvious from the definition of matching measures (Definition 5), as in this case $\mathcal{F} \cap \mathcal{G} = \mathcal{G}$ holds.

For property (2) let $C \in \mathcal{G}$ without loss of generality. Then we have

$$\mu(\mathcal{F}, \mathcal{G}) = \frac{m(\mathcal{F} \cap (\mathcal{G} - \{C\}))}{m(\mathcal{G} - \{C\}) + w(C)} \leq \frac{m(\mathcal{F} \cap (\mathcal{G} - \{C\}))}{m(\mathcal{G} - \{C\})} = \mu(\mathcal{F}, \mathcal{G} - \{C\}) \ .$$

For property (3) the case $C \notin \mathcal{F}$ is trivial. In case $C \in \mathcal{G} - \mathcal{F}$ holds, we get

$$\mu(\mathcal{F}, \mathcal{G}) = \frac{m(\mathcal{F} \cap \mathcal{G})}{m(\mathcal{G})} \leq \frac{m(\mathcal{F} \cap \mathcal{G}) + w(C)}{m(\mathcal{G})} = \mu(\mathcal{F} \cup \{C\}, \mathcal{G} \cup \{C\}) \ .$$

In case $C \notin \mathcal{G}$ first note that for any values a, b, c with $a \leq b$ we get $ab + ac \leq ab + bc$ and thus $\frac{a}{b} \leq \frac{a+c}{b+c}$. Thus, we get

$$\mu(\mathcal{F}, \mathcal{G}) = \frac{m(\mathcal{F} \cap \mathcal{G})}{m(\mathcal{G})} \leq \frac{m(\mathcal{F} \cap \mathcal{G}) + w(C)}{m(\mathcal{G}) + w(C)} = \mu(\mathcal{F} \cup \{C\}, \mathcal{G} \cup \{C\}) \ .$$

For property (4) the both sides of the equivalence is equivalent to $m(\mathcal{F}_1 \cap \mathcal{G}) > m(\mathcal{F}_2 \cap \mathcal{G})$, which completes te proof. □

Informally phrased property (1) states that whenever all requirements in a requested profile $\mathcal{G}$ (maybe even more) are satisfied by a given profile $\mathcal{F}$, then $\mathcal{F}$ is a perfect match for $\mathcal{G}$. Property (2) states that if a requirement not satisfied by a given profile $\mathcal{F}$ is removed from the requested profile $\mathcal{G}$, the given profile will become a better match for the restricted profile. Property (3) covers two cases. If $C \in \mathcal{G}$ holds, then simply the profile $\mathcal{F} \cup \{C\}$ satisfies more requirements than $\mathcal{F}$, so the matching value should increase. The case $C \notin \mathcal{G}$ is a bit more tricky, as the profile $\mathcal{G} \cup \{C\}$ contains an additional requirement, which is satisfied by the enlarged profile $\mathcal{F} \cup \{C\}$. In this case the matching value should increase, because the percentage of requirements that are satisfied increases. Property (4) states that if the given profile $\mathcal{F}_1$ is better suited for the required profile $\mathcal{G}$ than the given profile $\mathcal{F}_2$, then relative to $\mathcal{G}$ the profile $\mathcal{F}_2$ is less over-qualified than $\mathcal{F}_1$ for any other required profile $\mathcal{F}$, provided the intersections of $\mathcal{F} \cap \mathcal{G}$ with the two given profiles coincide.

Thus, disregarding for the moment our theory of matching measures, all four properties in Lemma 1 appear to be reasonable. Therefore, we require them as *plausibility constraints* that a human-defined mapping $h : \mathbb{F} \times \mathbb{F} \to [0, 1]$ should satisfy:

(1) $h(\mathcal{F}, \mathcal{G}) = 1$ for $\mathcal{G} \subseteq \mathcal{F}$,
(2) $h(\mathcal{F}, \mathcal{G}) \leq h(\mathcal{F}, \mathcal{G} - \{C\})$ for any concept $C \notin \mathcal{F}$, and
(3) $h(\mathcal{F}, \mathcal{G}) \leq h(\mathcal{F} \cup \{C\}, \mathcal{G} \cup \{C\})$ for any concept C.
(4) If $\mathcal{F} \cap \mathcal{F}_1 \cap \mathcal{G} = \mathcal{F} \cap \mathcal{F}_2 \cap \mathcal{G}$ holds, then

$$h(\mathcal{F}_1, \mathcal{G}) > h(\mathcal{F}_2, \mathcal{G}) \Leftrightarrow h(\mathcal{F}, \mathcal{F}_1 \cap \mathcal{G}) < h(\mathcal{F}, \mathcal{F}_2 \cap \mathcal{G}).$$

3.2 Derivation of Matching Measures

We want to show the following:

Theorem 1. *Let h be a human-defined matching measure that satisfies the plausibility constraints. Then there esists a matching measure μ that is ranking-preserving with respect to h.*

.

So let h be a human-defined matching measure that satisfies the plausibility constraints. Assume the lattice $\mathcal{L}$ contains $n+2$ elements $C_0, \dots, C_{n+1}$ with top- and bottom elements C_0 and C_{n+1}, respectively.

Fixing a requested profile $\mathcal{G}$, then $h(\mathcal{F}_1, \mathcal{G}) < h(\mathcal{F}_2, \mathcal{G})$ defines a linear inequation of the form $\sum\limits_{x \in U} x < \sum\limits_{x \in V} x$ with $U = \{w(C) \mid C \in \mathcal{F}_1 \cap \mathcal{G}\}$ and $V = \{w(C) \mid C \in \mathcal{F}_2 \cap \mathcal{G}\}$. In these inequalities we may remove summands $w(C)$ on both sides for $C \in \mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{G}$. In particular, $w(C_0)$ never appears in the inequalities. Without loss of generality we can also ignore C_{n+1}, as it only appears for the trivial case.

Analogously, if we fix a given profile $\mathcal{F}$, we also obtain linear inequalities of the form $\sum\limits_{x \in U} x < \sum\limits_{x \in V} x$ with disjoint sets U and V corresponding to sets of weights of concepts. If all these inequalities can be satisfied, then clearly the solution defines a matching measure μ that is order-preserving with respect to h.

The "worst case" arises, if we have a linear order on the set of all terms $\sum_{i \in I} x_i$, where x_i represents $w(C_i)$ and $I \subseteq \{1, \dots, n\}$. This arises for the lattice $\mathcal{L}$, in which all C_i $(i = 1, \dots, n)$ are pairwise incomparable. For all other lattices we can extend the set of inequalities derived from h to the "worst case".

So we reduce the problem of proving Theorem 1 to a problem of solving a set of linear inequalities. Thus, let $\mathcal{P}$ be a linear order on the set of terms $\{\sum_{i \in I} x_i \mid I \subseteq \{1, \dots, n\}\}$. We say that $\mathcal{P}$ is *realisable* iff there is a substitution $v : \{x_1, \dots, x_n\} \to \mathbb{R}^+$ of the variables by positive real numbers such that $\sum_{i \in I} x_i$ precedes $\sum_{j \in J} x_j$ in $\mathcal{P}$ iff $\sum_{i \in I} v(x_i) < \sum_{j \in J} v(x_j)$ holds. As all sums are finite, it is no loss of generality to seek substitutions by rational numbers, and further using the common denominator it suffices to consider positive integers only.

As $\mathcal{P}$ is defined by h we can assume that $\sum_{i \in I} x_i$ precedes $\sum_{j \in J} x_j$ for $I \subset J$. We can then also extend $\mathcal{P}$ to a partial order $\hat{\mathcal{P}}$ on multisets of variables by adding the same variable(s) to both sides. Clearly, $\mathcal{P}$ is realisable iff $\hat{\mathcal{P}}$ is realisable. For convenience we introduce the notation $U \prec V$ for multisets U, V over $\{x_1, \dots, x_n\}$ to denote the inequality $\sum x_i \in U m_U(x_i) x_i < \sum x_j \in V m_V(x_j) x_j$, where m_U and m_V are the multiplicities for the two multisets.

Theorem 2. *$\mathcal{P}$ is realisable iff there is no positive integer combination of inequalities in $\mathcal{P}$ that results in $A \prec B$ with $B \subseteq A$ as mutisets, i.e. $m_B(x_i) \leq m_A(x_i)$ for all $i = 1, \dots, n$.*

Proof (sketch). The necessity of the condition is obvious. For the sufficiency we use *Fourier-Motzkin elimination.* □

We now use Theorem 2 to prove our main result in Theorem 1.

Proof (of Theorem 1). Assume that $\mathcal{P}$ is not realisable. Then according to Theorem 2 there exist inequalities $U_1 < V_1, \dots, U_k < V_k$ in $\mathcal{P}$ such that $V = \biguplus i = 1^k V_i \subseteq \biguplus i = 1^k U_i = U$ as multisets and

$$\sum_{x \in U} \sum_{j=i}^{k} m_{U_j}(x)x < \sum_{x \in V} \sum_{j=i}^{k} m_{V_j}(x)x.$$

Let this system of inequalities be minimal, so each subset violates the condition in Theorem 2. Taking the inequalities in the some order let

$$A_i = \sum_{x} \sum_{j=1}^{i} m_{U_j}(x)x \qquad \text{and} \qquad B_i = \sum_{x} \sum_{j=1}^{i} m_{V_j}(x)x.$$

Then for $i < k$ there always exists some x with $\sum_{j=1}^{i} m_{U_j}(x) < \sum_{x} \sum_{j=1}^{i} m_{V_j}(x)$, while $A_i \prec B_i$.

On the other hand $\sum_{j=1}^{k} m_{U_j}(x) \geq \sum_{x} \sum_{j=1}^{k} m_{V_j}(x)$, while $A_k \prec B_k$.

Let V_i' be the multiset $B_i - A_i$, i.e. the multiset of all x with $m_{B_i}(x) > m_{A_i}(x)$ such that $m_{V_i'}(x) = m_{B_i}(x) - m_{A_i}(x)$ holds. Each $x \in V_i'$ is a witness for the violation of the condition in Theorem 2. In particular, we have $V_i' \neq \emptyset$ for all $i < k$, but $V_k' = \emptyset$.

Let $V_{i+1}'' = V_i \cap V_{i+1}$ as multisets, so $m_{V_{i+1}''}(x) = \min(m_{V_i'}(x), m_{V_{i+1}'}(x))$, i.e. x will at most be added to B_i to gibe B_{i+1}, but not to A_i. In particular, $V_{i+1}'' \subseteq V_i'$. Take the complement U_{i+1}' such that $V_i' = U_{i+1}' \uplus V_{i+1}''$.

As $U_i < V_i$ is in $\mathcal{P}$, we also have $U_i' < V_i$ in $\mathcal{P}$ for all $i > 1$ (U_1' is not yet defined).

Let $B_1' = V_1 = V_1'$. Then proceed inductively defining $W_i = B_i' - U_{i+1}'$ as well as

$$A_{i+1}' = U_{i+1}' \uplus W_i \qquad \text{and} \qquad B_{i+1}' = V_{i+1} \uplus W_i,$$

which gives $A_{i+1}' < B_{i+1}'$ in $\hat{\mathcal{P}}$ and $B_i' = A_{i+1}'$. That is, we obtain a chain

$$B_1' = A_2' < B_2' = \cdots < B_{k-1}' = A_k' < B_k'.$$

Complement these definitions by $U'_1 = B'_k \cap U_1 = A'_1$, and $X_0 = B'_k - U_1 = X_1$. Proceed inductively defining

$$\begin{aligned} C_1 = U'_1 \uplus X_0 \prec V_1 \uplus X_1 = B'_1 \uplus X_1 = D_1 \\ \text{and} \qquad X_{i+1} = X_i - (U_{i+1} - A'_{i+1}) \end{aligned}$$

This gives $C_{i+1} = A'_{i+1} \uplus X_i \prec B'_{i+1} \uplus X_{i+1} = D_{i+1}$ and $C_{i+1} = D_i$. Finally, due to this construction we also have $X_i \supseteq X_{i+1}$ for all i and $X_k = \emptyset$. This implies $D_k = C_1$, which means we have a cycle in $\hat{\mathcal{P}}$ contradicting the fact that it is a partial order. Therefore, $\mathcal{P}$ must be realisable. □

Example 4. (1) To illustrate the construction in the proof take the inequalities $x_1 + x_2 < x_3 + x_4$, $x_2 + x_3 < x_5$, $x_4 + x_5 < x_1 + x_3$ and $x_3 < x_2$ in $\mathcal{P}$ which satisfy the condition in Theoren 2. From these we construct first the following inequalities in $\hat{\mathcal{P}}$: $x_3 + x_4 < x_4 + x_5$, $x_4 + x_5 < x_1 + x_3$ and $x_1 + x_3 < x_1 + x_2$. As the last right hand side is U_1, we get $X_0 = \emptyset$, which defines the additional inequality $x_1 + x_2 < x_3 + x_4$. These four inequalities define the contradictory cycle in $\hat{\mathcal{P}}$.

(2) If the third inequality had been $x_4 + x_5 < x_1 + x_2 + x_3$ instead, the constructed inequalities in $\hat{\mathcal{P}}$ would have been $x_3 + x_4 < x_4 + x_5$, $x_4 + x_5 < x_1 + x_2 + x_3$ and $x_1 + x_2 + x_3 < x_1 + 2x_2$, this defines $X_0 = \{x_2\}$, which gives the addition inequality $x_1 + 2x_2 < x_2 + x_3 + x_4$ and thre modified inequality $x_2 + x_3 + x_4 < x_4 + x_5$, which again defines a cycle.

Note that our main result only states the existence of a ranking preserving matching measure μ. However, we obtain solutions for the linear inequations defined by h by minimising $x_1 + \cdots + x_n - 1$ under the conditions $\sum_{x_j \in V} x_j - \sum_{x_i \in U} x_i \leq \epsilon$ for a very small constant $\epsilon > 0$. For this linear optimisation problem the well-known *simplex algorithm* can be exploited.

4 Conclusion

In this paper we addressed the problem to determine matching measures for profiles that produce rankings, which are in accordance with the measures given by a human expert or at least imply the same rankings. For this we analysed linear inequalities that result from the human-defined rankings. We could show that if certain plausibility rules are obeyed by the human expert – i.e. the matching and the rankings are not biased by criteria not represented in the knowledge base – then we can indeed create such matching measures, with the help of which the human expertise can be approximated. This shows that the very general approach to matching based on filters provides the necessary flexibility required for diverse matching tasks.

This is only a starting point for even more sophisticated matching analysis aiming at consensus building among different experts and determination of the most suitable matching measure that is in accordance with the expert knowledge. We also have to take into account that valuations given by human experts will never be complete. This will be addressed in our future research.

References

1. F. Baader et al., editors. *The Description Logic Handbook: Theory, Implementation and Applications.* Cambridge University Press, 2003.
2. European distionary of skills and competences. http://www.disco-tools.eu.
3. T. Falk et al. Semantic-Web-Technologien in der Arbeitsplatzvermittlung. *Informatik Spektrum*, 29(3):201–209, 2006.
4. B. Ganter and C. Meschke. A formal concept analysis approach to rough data tables. *Transactions on Rough Sets*, 14:37–61, 2011.
5. B. Ganter, G. Stumme, and R. Wille. Formal concept analysis: Theory and applications. *Journal of Universal Computer Science*, 10(8):926, 2004.
6. B. Ganter and R. Wille. *Formal concept analysis - mathematical foundations.* Springer, 1999.
7. B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. F. Patel-Schneider, and U. Sattler. OWL 2: The next step for OWL. *Journal of Web Semantics*, 6(4):309–322, 2008.
8. International standard classification of education. http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx.
9. International standard classification of occupations. http://www.ilo.org/public/english/bureau/stat/isco/isco08/, 2008.
10. M. Levandowsky and D. Winter. Distance between sets. *Nature*, 234(5):34–35, 1971.
11. D. Looser, H. Ma, and K.-D. Schewe. Using formal concept analysis for ontology maintenance in human resource recruitment. In F. Ferrarotti and G. Grossmann, editors, *Ninth Asia-Pacific Conference on Conceptual Modelling (APCCM 2013)*, volume 143 of *CRPIT*, pages 61–68. Australian Computer Society, 2013.
12. M. Mochol, H. Wache, and L. J. B. Nixon. Improving the accuracy of job search with semantic techniques. In W. Abramowicz, editor, *Business Information Systems, 10th International Conference (BIS 2007)*, volume 4439 of *Lecture Notes in Computer Science*, pages 301–313. Springer, 2007.
13. L. Paoletti, J. Martinez-Gil, and K.-D. Schewe. Extending knowledge-based profile matching in the human resources domain. submitted for publication, 2015.
14. N. Popov and T. Jebelean. Semantic matching for job search engines – a logical approach. Technical Report 13-02, Research Institute for Symbolic Computation, JKU Linz, 2013.