Experiences with Modeling Composite Phenotypes in the SKELETOME Project

Tudor Groza[1], Andreas Zankl[2], and Jane Hunter[1]

[1] School of ITEE, The University of Queensland, Australia
tudor.groza@uq.edu.au, jane@itee.uq.edu.au
[2] Bone Dysplasia Research Group
UQ Centre for Clinical Research (UQCCR)
The University of Queensland, Australia
a.zankl@uq.edu.au

Abstract. Semantic annotation of patient data in the skeletal dysplasia domain (e.g., clinical summaries) is a challenging process due to the structural and lexical differences existing between the terms used to describe radiographic findings. In this paper we propose an ontology aimed at representing the intrinsic structure of such radiographic findings in a standard manner, in order to bridge the different lexical variations of the actual terms. Furthermore, we describe and evaluate an algorithm capable of mapping concepts of this ontology to exact or broader terms in the main phenotype ontology used in the bone dysplasia domain.

1 Introduction

Skeletal dysplasias represent a group of rare genetic disorders affecting the skeletal development. Patients with such disorders suffer from complex medical issues that can be grouped into three categories: (i) clinical findings, i.e., pains in limbs; (ii) radiographic findings, i.e., bilateral arachnodactyly; (iii) genetic findings, i.e., deletion mutation in FGFR3. In a previous paper [1], we have introduced the SKELETOME project that has developed a community-driven knowledge curation platform for this domain, able to capture and integrate such clinical, radiographic and genetic findings. The underlying foundation of the platform is an ontology-driven knowledge engineering cycle introduced to bridge the current knowledge about the domain and the continuously growing pool of patient cases. The cycle has two phases: (1) semantic annotation – bridging knowledge to cases – and (2) collaborative diagnosis, collaborative knowledge curation and evolution – from cases to knowledge. The semantic annotation process relies on clinical and radiographic findings grounded in Human Phenotype Ontology (HPO) [2] concepts – one of the only phenotype ontologies for rare disorders.

In this paper we focus on issues associated with this semantic annotation process, and more precisely on representing, in a standard manner, radiographic findings present in X-Ray descriptions and clinical summaries. At the same time, since the phenotype knowledge in SKELETOME is modeled only via HPO concepts, we aim to map where possible, instances of this standard representation to

P. Cudré-Mauroux et al. (Eds.): ISWC 2012, Part II, LNCS 7650, pp. 82–97, 2012.

terms existing in HPO. The root of our problem lies in the structural and lexical differences that exist between the terms describing such radiographic findings.

There currently are two major "*vocabularies*" [1] used within the community: (i) a generic one, and (ii) the International Skeletal Dysplasia Registry (ISDR) vocabulary. The generic vocabulary, used by the vast majority of clinicians, consists of an unstructured and virtually unlimited set of terms representing associations between qualities and anatomical entities – entries in free text clinical summaries. The terms used to describe a patient case are subject to the personal style of the clinician documenting the case, and hence may take different granularities and lexical groundings. For example, a clinician may use the term *bowed tibial shaft*, while another may use the term *angulation of the tibial diaphysis* to denote, in practice, the same thing. On the opposite side, the ISDR vocabulary, used at a much smaller scale only by ISDR, has a fixed and hierarchical set of around 270 terms representing anatomical parts, each having associated, in average, 5 to 10 qualities (hence a total of around 2,000 terms). Patient case findings will always have assigned terms from this set and each term will have the exact same structure and lexical grounding in all cases. For example, the corresponding entry for the two terms mentioned above would be: *Tibia – Diaphysis – Abnormality: Angulated*. The differences between the different lexical groundings of terms make the semantic annotation process very challenging.

In addition to the issues listed above, while HPO is, to date, the most comprehensive phenotype ontology for rare disorders, unfortunately, it is far from being complete. As a result, in order to provide a proper context for radiographic findings found in patient cases, we do not only require a mapping to existing HPO terms – where these exist, but also a mapping to the most appropriate parent within HPO – for those that don't have an exact match.

The contribution brought by this paper is two-fold: (i) we describe an ontology, the Phenotype Fragment Ontology (PFO), aimed at providing a standard structure for radiographic findings, independently of the actual lexical groundings, and (ii) we propose an algorithm that maps concepts modeled with this ontology to HPO terms by considering both exact and broader matches.

The goal of PFO is to capture the inner structure of radiographic findings by enabling the construction of complex phenotypes via combinations of anatomical entities (i.e., *Diaphysis* – partOf – *Tibia*) and qualities (i.e., *Bowed*). In practice, PFO provides a meta-model for phenotypes where the actual concepts (i.e., anatomical entities and qualities) are defined via well-known and widely adopted ontologies in the biomedical domain, such as the Foundational Model of Anatomy (FMA) [3] and the Phenotype and Trait Ontology (PATO) [4]. The granularity proposed by PFO does not only provide a solution to the issues discussed in this paper, but also enables a fine-grained exploration of the phenotype space in the bone dysplasia domain. This, in turn, enables the exploration of commonalities

[1] Throughout the paper, we use the term *vocabulary* to denote the structural and lexical commonalities that group a set of terms used to describe radiographic findings. From a semiotic and medical perspective the community uses a single set of terms.

between disorders based on the anatomical localization of phenotypes and the development of anatomical localization - oriented decision support methods.

Starting from concepts represented using PFO, we have developed a mechanism that maps them to exact or broader HPO terms. Since this mapping is part of the semantic annotation process in SKELETOME, the user plays a central role by validating the mapping results. Our algorithm provides a ranked list of candidate HPO terms, of which the top 5 are being presented to the user. Consequently, our focus has been on achieving a high precision.

Ontology matching (OM) has been a very active research area during the last decade. Systems like Falcon [5], RiMOM [6], SAMBO [7] or DSSim [8] have achieved impressive results during several OM challenges (see [9] for a comprehensive overview on OM). Apart from the lexical similarity performed by these systems (which is dependent on the lexical groundings), the overall mapping process we require is different. Hence, we were unable to directly use or compare against any of them. Ontology matching assumes the mapping of concepts from one ontology to corresponding concepts in another ontology. In our case, PFO does not provide actual concepts that could be directly mapped. It only provides the scaffolding onto which concepts can be created, while the actual semantics is provided by terms from FMA and PATO, used to compose PFO concepts [2].

The remainder of the paper is organized as follows. Section 2 provides a brief overview of HPO and discusses the motivation behind PFO. Section 3 describes PFO and its associated engineering process. In Section 4 we detail the mapping algorithm, and before concluding in Section 6, we discuss some experimental results and the shortfalls of our approach in Section 5.

2 Background and Motivation

2.1 Human Phenotype Ontology

HPO has been developed to provide a controlled vocabulary for phenotypic features encountered, in principle, in hereditary diseases listed in the Online Mendelian Inheritance in Man (OMIM) database [3]. The ontology, currently comprising around 9,900 concepts, describes three main streams [4]: (i) Mode of Inheritance, (ii) Onset and clinical course, and (iii) Phenotypic abnormalities.

Phenotypic abnormalities (the concepts of interest in our study) represent more than 95% of the ontology and are organized in a hierarchical manner (via class–subclass relations). This hierarchy is, in principle, based on the main anatomical systems, such as, the nervous system (`HP_0000707` – *Abnormality of the nervous system*) or the skeletal system (`HP_0000924` – *Abnormality of the*

[2] Mapping HPO terms to concepts from other phenotype ontologies has been previously discussed in the literature. However this also falls under Ontology Matching since the goal is to match `Cardiomegaly` from HPO, for example, to `Enlarged heart` from the Mammalian Phenotype Ontology.

[3] `http://www.omim.org/`

[4] Please note that all experiments discussed in this paper have been conducted on the HPO version from 31 May 2012.

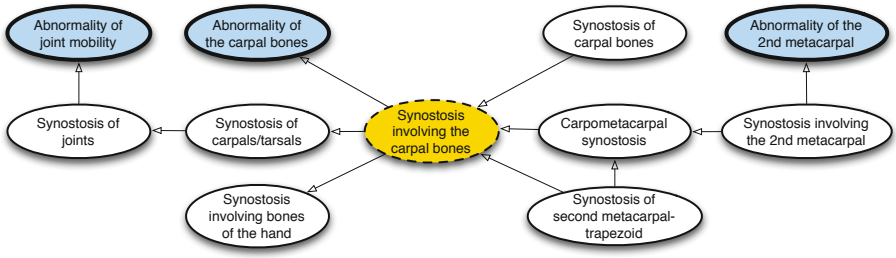

Fig. 1. A snapshot of the HPO structure (arrows denote class–subclass relations)

skeletal system). Each concept has a label and may have a definition and exact or related synonyms.

There are two aspects that raise challenges when using HPO: multiple inheritance and an overall inconsistency of the lexical representations of the concepts. The latter refers to the inconsistency in the inner lexical structure of the terms (i.e., *Synostosis of joints* vs. *Carpometacarpal synostosis*), as well as to the reuse of the same conceptual lexical grounding in multiple terms (e.g., `HP_0005048` – *Synostosis of carpal bones* has listed *Fusion of carpal bones* as synonym, while `HP_0009702` – *Synostosis involving the carpal bones* has listed as synonyms *Fused carpal bones*). Finally, a third issue is the use of generic lexical representations as synonyms for very specific concepts (e.g., `HP_0010239` – *Aplasia of the middle phalanges of the hand* has listed as synonym *Absent middle phalanges*). All these issues make the mapping process more complex and affect its resulting precision.

The multiple inheritance, on the other hand, requires one to adopt a goal-oriented interpretation of the hierarchy, if relying on the subclass relations between terms. Fig. 1 depicts an example of multiple inheritance extracted from HPO. This shows how at different levels in the hierarchy one may find concepts that are subclasses of both concepts defined based on the anatomical localization of the abnormality (i.e., *Abnormality of joint mobility*), as well as concepts defined based on the type of the abnormality (i.e., *Synostosis of carpals/tarsals*). In addition, the structure also contains relations that do not follow any of the two directions, e.g., the subclass relation between *Synostosis of second metacarpal-trapezoid* and *Synostosis involving the carpal bones*. In our case, these aspects have influenced the design of the mapping algorithm described in Section 4. More concretely, when defining broader matches between PFO and HPO concepts we have considered a correct match to be the most specific HPO ancestor that is defined based on the anatomical localization. For example, in this interpretation, *Synostosis involving the 2nd metacarpal* would have the closest broader match *Abnormality of the 2nd metacarpal*.

2.2 Analysis of Radiographic Findings

In order to gain a deeper understanding in the inner structure of the radiographic findings we have collected a list of 675 random findings from patient cases listed

Table 1. Coverage of our radiographic findings in HPO

Category	Total terms	Exact HPO match	Broader HPO match	No HPO match
Simple	387	237 (61%)	136 (35%)	14 (4%)
Composite anatomy	156	66 (42%)	56 (36%)	34 (22%)

in the European Skeletal Dysplasia Network registry and from a widely adopted text book in the bone dysplasia domain – Spranger et al., *Bone dysplasia: an atlas of genetic disorders of skeletal development* [5]. All items in this list can be categorised under the generic vocabulary introduced in Section 1. The nature of the ISDR vocabulary provided us with access to all possible terms in it, hence there was no need for a collection process. Also, the analysis focuses only on the generic vocabulary since this is the one to introduce the major issues previously described. The size of the list represents around a fifth of the total number of findings present under the HPO *Abnormality of the skeletal system* (3,744 sub-concepts), which is of particular interest for the skeletal dysplasia domain.

The analysis of the list of radiographic findings has revealed that they can be grouped, based on their inner structure, into three categories: (i) simple findings, i.e., associations between a single anatomical entity and qualities – *flat skull* (387 findings); (ii) composite anatomy findings, i.e., associations of composed anatomic entities and qualities – *bifid distal phalanx of the thumb* (156 findings); and (iii) composite phenotypes, i.e., conjunctions of findings from the previous two categories – *curved femora with rounded distal epiphyses* (132 findings).

As a next step, we investigated to what extent is HPO able to cover these findings, with a focus only on the first two categories as findings in the last category are covered by combining existing findings in the first two. Consequently, we have manually mapped the findings in the list to HPO terms. Table 1 summarises the mapping results. In the "simple" category 237 findings (61%) had an exact match, 136 findings (35%) could be associated with a broader match (considering the interpretation for broader matches given in the previous section), while the rest of 14 (4%) could not be mapped. Two reasons made the mapping impossible for this last set: the findings represented *normal* states (i.e., *normal pelvis*), which have no correspondence in HPO (since it models only abnormal findings) or the findings were too generic and their interpretation was subject to a particular context (i.e., *gracile bones* – in the context of a particular X-Ray, *bones* could refer to, for example, *phalanges*). Similarly, in the "composite anatomy" category 66 findings (42%) had an exact match, 56 (36%) a broader match, and 34 (22%) we were unable to map.

The exact match manual mappings have then been used as ground truth in an attempt to automatically map the findings to HPO terms. In this experiment, we've used the NCBO Annotator [10] and three well-known string similarity measures: Levenstein, Needleman-Wunch and Smith-Waterman (the last two are

[5] The list can be downloaded from: `http://tiny.cc/grtifw`

Table 2. String similarity measures performance on the radiographic findings

Similarity	P@1	P@2	P@3	P@4	P@5
Category: Simple					
Levenstein	**0.55**	0.29	0.21	0.16	0.13
N-W	0.5	0.27	0.19	0.14	0.11
S-W	0.54	0.29	0.2	0.15	0.13
Category: Composite anatomy					
Levenstein	**0.41**	0.25	0.17	0.13	0.11
N-W	0.38	0.21	0.14	0.11	0.09
S-W	0.36	0.24	0.17	0.14	0.12

Table 3. NCBO Annotator annotation performance on the radiographic findings

Category	Precision	Recall	F1 score
Simple	0.79	0.37	0.5
Composite anatomy	0.57	0.2	0.3

the reference algorithms used to align gene/protein sequences in Bioinformatics). In order to ensure a fair comparison, in the case of the similarity measures, the 1-to-1 mapping (finding – HPO term) has been realized by computing the similarity of the finding against the label and all synonyms of the HPO term. The highest similarity score was returned as the final similarity between the two. We've evaluated the resulting performances by looking at precision at k (P@k) with k=1, 2, 3, 4, 5 as presented in Table 2. In the case of the NCBO Annotator, since the annotation results are dichotomous (i.e., a match is either found or not) we have calculated the standard precision, recall and F1 (results are listed in Table 3).

The result of this experiment shows (if it was necessary) that the mapping process is very sensitive to the lexical representation of the findings. As an additional remark, the example provided in the introduction (i.e., *bowed tibial shaft* vs. *angulation of the tibial diaphysis*) is a typical case that can make the difference between a mapping hit and a miss. Overall, the NCBO Annotator had a very good precision at the expense of the recall, while among the similarity measures the best performance has been achieved by the Levenstein distance in both categories of findings. The two results are obviously not directly comparable, but they did provide us with a good overview of what we can achieve, with respect to our goal, with off-the-shelf solutions.

Consequently, we saw the need to create a standard format for these findings, abstracting from the actual lexical representation (i.e., capturing the *object* as opposed to the *symbol*, from a semiotical perspective), and for which we can design a generic mapping mechanism. This standard structure is provided by our Phenotype Fragment Ontology (PFO), described next. Another obvious solution

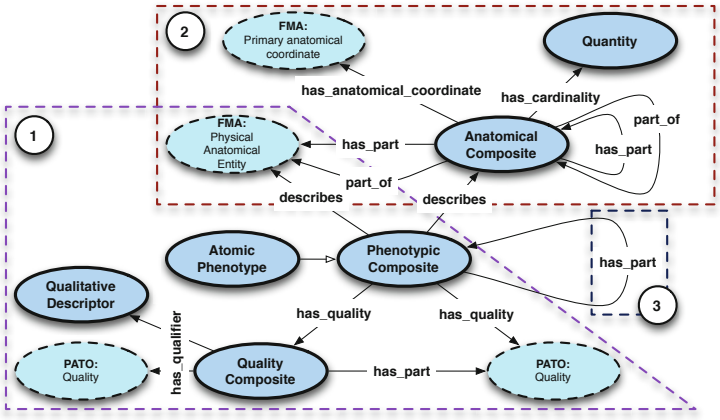

Fig. 2. Snapshot of the main part of the Phenotype Fragment Ontology

would have been to design a direct mapping algorithm between findings and HPO terms, sensitive to the actual terms yet, probably, achieving a fairly high accuracy. However, this would have served solely its design purpose, i.e., the direct mapping. PFO, on the other hand, opens several new research paths by enabling the exploration of radiographic findings at a level never achieved before.

3 The Phenotype Fragment Ontology

The Phenotype Fragment Ontology (PFO) has two main goals: (i) to provide a standard representation for radiographic findings, based on their intrinsic structure, and (ii) to enable the creation of the corresponding concepts by re-using concepts from widely adopted ontologies, i.e., FMA and PATO.

Fig. 2 depicts the core part of PFO [6] together with the design steps. The central concept of the ontology, the `Phenotypic Composite`, carries a bridging role between the anatomical side of the findings and the qualities they bear. Starting from this central concept, the design of PFO has followed closely the result of the experiment detailed in Section 2. As a first step (no. 1 in the figure) we have added support for modeling simple findings, as associations of `FMA: Physical Anatomical Entity` (via the *describes* relation) and qualities (via the *has_quality* relation), which can either be a `PATO: Quality` or a `Quality Composite`. The latter can then be further expressed as an association between an explicit quality and a qualifier. Most qualities and qualifiers are directly reused from PATO, however, where this is not possible, PFO introduces its own concepts. Secondly, we have introduced the `Anatomical Composite` concept to enable the modeling of the second category of findings. This, may have (*has_part*) or be a *part of* a `FMA: Physical Anatomical Entity` and may have attached

[6] `http://purl.org/skeletome/phenotype`

an anatomical coordinate (the figure only presents one such type of anatomical coordinate, but the ontology contains several) or a cardinality. Finally (no. 3 in the figure), in order to capture the third category (i.e., composite phenotypes per se), we have added the *has_part* self-relation on `Phenotypic Composite`.

Returning to the example introduced in Section 1, it can be observed that, independently of the lexical grounding (*bowed tibial shaft* or *angulation of the tibial diaphysis*), the concept denoting the abstract radiographic finding will have the following structure: a `Phenotype Composite` that

- *describes* an `Anatomical Composite`, which *has_part* `FMA: Diaphysis`, and is *part_of* `FMA: Tibia`, and
- *has_quality* `PATO: PATO_0000406` (*Bowed*).

At the same time, this structure maps perfectly onto the structure of the ISDR concepts. Similarly, *Flat skull* is modeled as: `Phenotype Composite` *describes* `FMA: Skull` and *has_quality* `PATO: PATO_0000407` (*flat*). For conciseness we have not included the logical definitions of these concepts. However, the definitions of all concepts discussed in the previous section, from the simple and composite anatomy findings categories, plus some examples from the composite phenotypes category, can be found at `http://purl.org/skeletome/spo`. These definitions have been also used for the evaluation described in Section 5.

While lightweight, PFO enables the definition of standard representations for very rich radiographic findings. In addition, it brings a series of side advantages, such as: (i) a **standard lexicon**, since the lexical grounding of the concepts will always be provided by the ontologies that underpin its definition; (ii) explicit modeling of **cardinality**, which is important from a clinical perspective; (iii) the possibility of modeling **normal** phenotypes, again important from a clinical and decisional perspective; and (iv) the scaffolding for decomposing **atomic elements**, i.e., monolithic terms that do not reveal in their lexical representation the association anatomical localization – quality; for example, *macrocephaly*, which can be represented as `FMA: Head` – `PATO: PATO_0000586` (*Large*)

On the negative side, relying on external ontologies introduces a series of issues, one of which is the lack of concepts to represent certain anatomical parts or qualities (the rest are discussed in Section 5). For example, FMA has no corresponding concept for *Müllerian duct*, while PATO does not cover most of the "metaphoric" qualities, such as *angel-shaped* or *cloverleaf* – which we had to introduce in PFO.

4 Mapping PFO Concepts to HPO Terms

As mentioned in Section 1 our second goal is to map radiographic findings modelled as PFO concepts to exact or broader terms in HPO. Consequently, we have developed a mapping algorithm that ranks all HPO terms according to their similarity to a given PFO concept (described in Section 4.1). This algorithm has been evaluated for exact matching in Section 5. For finding broader HPO terms, we've added a series of extra steps to the general algorithm, as detailed in Section 4.4.

4.1 General Mapping Algorithm

Algorithm 1 lists the general mapping algorithm, while snippets 2, 3 and 4 present some specific methods used in it. In order to get a better understanding of the algorithm we will consider as example, the 1-to-1 mapping of `HP_0009611` (*Notched terminal thumb phalanx*) to `Bifid_distal_phalanx_of_the_thumb` [7].

Lexicon Creation and Tokenization. A PFO concept has two main elements, i.e., the anatomical part and the quality, each of which may be associated with an extra set of elements, i.e., an anatomical coordinate and qualifiers – as described in Section 3. For each of these four elements, we generate the corresponding lexicons using the concepts that underpin their definition. Each lexicon comprises the label and all the synonyms of the given concept. Subsequently, all entries in the lexicons are tokenized. The same procedure is also employed on the HPO concept. In our example, the PFO concept has two anatomical entities (`FMA: Phalanx of finger` and `FMA: Thumb`), one anatomical coordinate (`Distal`) and one quality (`Bifid`). Hence, the result of this step is:

- $LexAnat_C\_TOKENS$ = { [Phalanx of finger ← (phalanx, of, finger), Hand phalanx ← (hand, phalanx), ...], [Thumb ← (thumb), First digit of hand ← (first, digit, of, hand)] }
- $LexCoord_C\_TOKENS$ = { [Distal ← (distal), Terminal ← (terminal)]}
- $LexQual_C\_TOKENS$ = { [Bifid ← (bifid), Forked ← (forked)]}

Similarly, the result for the HPO concept is:

- HPO_C_TOKENS = { [Notched terminal thumb phalanx ← (notched, terminal, thumb, phalanx), Bifid distal phalanx of thumb ← (bifid, distal, phalanx, of, thumb]), ...] }

Similarity Matrix and Traces Computation. For each entry in each lexicon of the PFO concept (i.e., Anat, Coord, Qual and Qualif) we compute a similarity matrix and associated traces against each entry in the lexicon of the HPO concept. As a remark, we use the term *trace* to denote the maximal diagonal in a similarity matrix and not the usual trace that can be computed only in squared matrices. Section 4.2 details this process and exemplifies it for *Phalanx of finger* vs. *Bifid distal phalanx of thumb*. For each entry in each lexicon of the PFO concept this step results in a list of associations (HPO lexicon entry – max trace) for both the full and the optimal length of the entry. As an example, *Phalanx of finger* will have the following (partial) result:

- Full length traces: {(notched terminal thumb phalanx – 0.33), Length: 3; (bifid distal phalanx of thumb – 0.66), Length: 3}
- Optimal length traces: {(notched terminal thumb phalanx – 0.59), Length: 1 (only *phalanx* is used); (bifid distal phalanx of thumb – 0.66), Length: 3}

[7] `http://purl.org/skeletome/spo#Bifid_distal_phalanx_of_the_thumb`

Algorithm 1. General mapping algorithm

Require: PFO_C

$PFO_C = \{Anat_C, Coord_C, Qual_C, Qualif_C\}$
$LEX_C = \{LexAnat_C, LexCoord_C, LexQual_C, LexQualif_C\}$,
where $LexX_C = \{label, syn_1, \ldots, syn_n\}$ of X_C and X ∈ {Anat, Coord, Qual, Qualif}

// Tokenization of all lexical groundings of each entry in a particular lexicon
for all $LexX_C \in LEX_C$ **do**
    $LexX_C\_TOKENS = \{LexX_C\_Tokens_i = [t_1, t_2, \ldots, t_N], i = [1, N]\}$
    where N = No. synonyms + 1 (the label) and $X \in \{Anat, Coord, Qual, Qualif\}$
end for

for all $HPO_C \in HPO$ **do**
    // Tokenization of all lexical groundings of the HPO concept
    $HPO_C\_TOKENS = \{HPO_C\_ENTRY_i = [t_1, t_2, \ldots, t_N], i = [1, N]\}$
    for all $X \in \{Anat, Coord, Qual, Qualif\}$ **do**
        for i:= 1 to N **do**
            Consider $LexX_C\_TOKENS_i$ // E.g., $LexAnat_C\_TOKENS_i$
            $SimX_i = similarity\_matrix\_and\_traces(LexX_C\_TOKENS_i, HPO_C\_TOKENS)$
            $OptSimX_i = length\_based\_optimal\_traces(SimX_i, HPO_C\_TOKENS)$
        end for
    end for

    SimAnat = $\{SimAnat_N, \ldots, SimAnat_1\}$
    FullSimAnat = full_anat_traces(HPO_C_TOKENS, SimAnat)

    $HPO_C\_SIM_i$ = aggregated_similarity(OptSimX, FullSimAnat), X ∈ {Coord, Qual, Qualif}
    // See Section 4.3 for the aggregated similarity computation
    $HPO_C\_SIM = max\|HPO_C\_SIM_i\|$
end for

Algorithm 2. Similarity matrix and traces

Require: $LexX_C\_TOKENS_i, HPO_C\_TOKENS$

TRACES = {}
for j:= 1 to N **do**
    SIM_MAT_j = similarity_matrix($LexX_C\_TOKENS_i, HPO_C\_ENTRY_j$)
    // See Section 4.2 for the similarity matrix and traces computation

    Trace($HPO_C\_TOKENS_j$) = compute_traces(SIM_MAT_j)
    // Trace($HPO_C\_TOKENS_j$) = {Value, Length, Start_Index}

    TRACES = TRACES ∪ $\{HPO_C\_ENTRY_j \rightarrow Trace(HPO_C\_TOKENS_j)\}$
end for

return TRACES

Length-Based Optimal Traces. For each entry in each lexicon of the PFO concept, this step reduces the list of associations produced by the previous one by choosing the maximal trace for a particular length. Continuing the example above, for *Phalanx of finger* this step will produce:

- Length:3 ← (bifid distal phalanx of thumb – 0.66)
- Length:1 ← (notched terminal thumb phalanx – 0.59)

Full Anatomy Traces. Until this point each element of the PFO concept has been considered individually, including the different anatomical parts. In our example, we have calculated the length-based optimal traces for both *Phalanx of finger* (and the rest of its lexicon entries), as well as for *Thumb* (and the rest of its lexicon entries). This step reunites all anatomical parts by looking for the optimal full anatomy trace for each HPO lexicon entry. This is done by averaging the individual anatomical traces for a particular HPO lexicon entry and then choosing the trace with the highest score. For example, if for the HPO lexicon entry *bifid distal phalanx of thumb* we have the following:

- 1: (Phalanx of finger – 0.59), (Thumb – 0.99) → 0.79
- 2: (Hand phalanx – 0.34), (Thumb – 0.99) → 0.66

this step will choose option 1 has being the optimal full anatomy trace.

Similarity Aggregration. Taking the length based optimal traces and the full anatomy traces produced above, this step computes the final similarity as described in Section 4.3.

Algorithm 3. Length based optimal traces

Require: $SimX_i, HPO_C\_TOKENS$
1: LENGTH_BASED_OPT = {}
2: **for all** $HPO_C\_ENTRY_j \in SimX_i$ **do**
3: Trace ∈ LENGTH_BASED_OPT
4: **if** Trace.Length == $Trace(HPO_C\_TOKENS_j.Length$ **then**
5: LENGTH_BASED_OPT = LENGTH_BASED_OPT ∪ $max\|Trace(HPO_C\_TOKENS_j.Value, Trace.Value\|$
6: **end if**
7: **end for**
8: return LENGTH_BASED_OPT

Algorithm 4. Full anatomy traces

Require: HPO_C_TOKENS, SimAnat
1: FULL_SIM = {}
2: **for all** $HPO_C\_ENTRY_j$ **do**
3: OPT_FULL_SIM = $max\|\frac{1}{N} * \sum_{i=1}^{N} SimAnat_{ij}\|$
4: FULL_SIM = FULL_SIM ∪ $\{HPO_C\_ENTRY_j \rightarrow OPT\_FULL\_SIM\}$
5: **end for**
6: return FULL_SIM

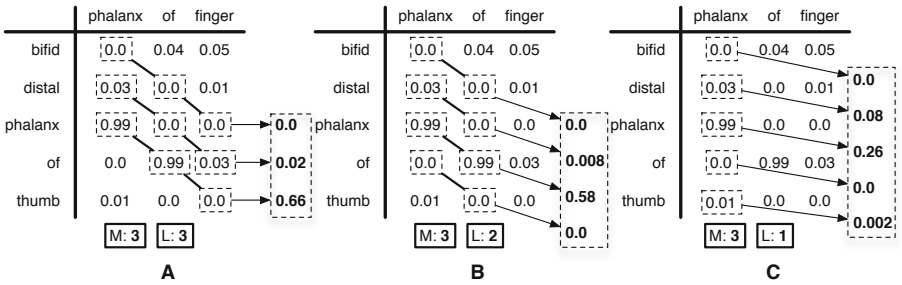

Fig. 3. Example of similarity matrix and traces computation for L = 3, 2, 1

4.2 Similarity Matrix and Traces Computation

The similarity matrix and traces computation is always performed on pair of tokenized lexicon entries, e.g., a lexicon entry for an anatomical part such as *Phalanx of finger* and a HPO concept lexicon entry, such as *bifid distal phalanx of thumb*. The goal of this step is to find the segment in the HPO concept lexicon entry that best matches the lexicon entry of, in our example, the anatomical part for different lengths of this last lexicon entry. We start by creating a M x N similarity matrix, where M is the length of the anatomical part and N is the length of the HPO lexicon entry. Fig. 3 depicts a full example of this step using the above pair of lexicon entries. The values in the similarity matrix are given by the following string similarity metric:

$$sim(s_1, s_2) = w_1 * sim_1(s_1, s_2) + w_2 * sim_2(s_1, s_2) + w_3 * sim_3(s_1, s_2) \quad (1)$$

where sim_1, sim_2 and sim_3 are defined below and correspond to the normalized longest common subsequence (LCS) and the normalized maximal consecutive longest common subsequence (MCLCS) starting at 1 (i.e., with the first character) and respectively at n (i.e., starting anywhere in the string). The weights w_1, w_2 and w_3 have been set empirically to 0.3, 0.6 and 0.1, because in the case of anatomical parts their prefix and root provide a higher accuracy in similarity matching (e.g.., *tibia*, *tibiae*, *tibial*).

$$sim_1 = NLCS(s_1, s_2) = \frac{length(LCS(s_1, s_2))^2}{length(s_1) * length(s_2)} \quad (2)$$

$$sim_2 = NMCLCS_1(s_1, s_2) = \frac{length(MCLCS_1(s_1, s_2))^2}{length(s_1) * length(s_2)} \quad (3)$$

$$sim_3 = NMCLCS_n(s_1, s_2) = \frac{length(MCLCS_n(s_1, s_2))^2}{length(s_1) * length(s_2)} \quad (4)$$

The trace of the similarity matrix is the multiplication of the arithmetic mean of the matrix diagonal for a given length L $\leq$ M with a penalty factor, as per Eq. 5. The penalty factor (the first component of Eq. 5.) is a monotonically decreasing

function that penalises the trace for all the non-stop word tokens omitted from the initial lexicon entry (i.e., M - L). For example, part C of Fig. 3 shows the trace computation for L = 1, i.e., the token *Phalanx*. In this case, the penalty factor depends on M - L = 1 (*finger*), since the token *of* is a stop word.

$$Trace(SimMat) = (e^{-\frac{M-L}{M}} - \frac{M-L}{M*e}) * \frac{\sum_{i=1}^{L} SimMat_{ii}}{L} \quad (5)$$

4.3 Similarity Aggregation

The overall similarity is the paired aggregation of the similarities of each of the four elements of the PFO concept, i.e., Anat – Coord and Qual – Qualif. We consider them in pairs because the anatomical coordinate is an extension of the anatomical part, and hence it is directly dependent on it, while the qualifier is an extension of the quality. Eq. 6 shows the overall similarity, with sim_{A-C} (i.e., the joint Anat – Coord similarity) being expressed in Eq. 7 and sim_{Q-Q} (i.e., the joint Qual – Qualif similarity) being expressed in Eq. 8. The two components of the overall similarity are: (i) a penalty factor depending on t_L – the number of HPO tokens left out from the similarity computation, and is the same as in the similarity matrix and trace calculation; and (ii) the aggregation of the two above mentioned similarities that gives more weight to the anatomical similarity.

The individual Anat – Coord similarity is computed by raising the multiplication of the final Coord trace with the arithmetic mean of all Anat traces to the power of e – the higher the multiplication score → the higher the similarity. Finally, the individual Qual – Qualif similarity is the arithmetic mean of the Qualif trace and the arithmetic mean of all Qual traces.

$$sim(PFO, HPO) = (e^{-t_L} - \frac{t_L}{e}) * \frac{6 * sim_{A-C} * sim_{Q-Q}}{2 * (2 * sim_{A-C} + sim_{Q-Q})} \quad (6)$$

$$sim_{A-C} = (\frac{TraceCoord}{N} * \sum_{i=1}^{N} TraceAnat_i)^e \quad (7)$$

$$sim_{Q-Q} = \frac{1}{2} * (TraceQualif + \sum_{i=1}^{N} TraceQual_i) \quad (8)$$

4.4 Broader Mapping Algorithm

The broader mapping algorithm extends the general one with two more steps. Firstly, it generates the ranked list of similarities on all HPO concepts using the general mapping algorithm and retains only those candidates that have the maximum similarity. Secondly, for each pair of candidates in the filtered list, it looks for the lowest common ancestor (LCA) from HPO and maps the LCA to the list of corresponding candidates. It then computes the standard deviation of the sizes of the candidates list associated with each LCA and retains only

Table 4. Evaluation results of the mapping process

Category	P@1	P@2	P@3	P@4	P@5
Exact match					
Simple	**0.85**	0.49	0.35	0.28	0.24
Composite anatomy	**0.75**	0.48	0.39	0.32	0.28
Broader match					
Simple	**0.91**	0.50	0.37	0.30	0.25
Composite Anatomy	**0.72**	0.46	0.36	0.31	0.27

those LCAs that have their corresponding list size greater than the standard deviation. Finally, these LCAs are used as input for another general mapping run, against the original PFO concept, however, this time by using only the anatomical and anatomical coordinates similarities. This last mapping is driven by the interpretation of the HPO hierarchy we have introduced in Section 2 in which the classification is done based on the anatomical localization, and hence the quality of the broader concept should be broader than the quality of the PFO concept under scrutiny.

5 Experimental Results

We have evaluated both mapping algorithms on all the concepts in the simple and composite anatomy category mentioned in Section 2. As already mentioned, their logical definitions can be found at `http://purl.org/skeletome/spo`. Similarly to the experiment described in Section 2, and following the goal set in Section 1. we have looked at precision at k (P@k) with k=1, 2, 3, 4, 5 (see Table 4).

The exact mapping of simple findings has achieved a maximal P@1 of 0.85, while the best exact matching on composite anatomy findings has been 0.75 at P@1. In both experiments (i.e., exact and broader) the composite anatomy has achieved lower precision results due of increased number of false positives it may introduce for each of the composing anatomical parts. A careful analysis of the missed mappings has revealed the following aspects:

1. Most of the failed mappings, especially in the composite anatomy category, are due to the HPO inconsistencies at the lexical representation level (as mentioned in Section 2). More concretely, three aspects have caused issues: (1) inconsistencies in using proper quality terms – i.e., using terms such as *hypoplastic* and *short / small*, or *aplastic* and *absent* in an alternative manner, although such terms have a clear individual semantics; (2) ambiguous quality definitions – i.e., *hypoplasia / small* as a quality of a finding; (3) the presence of generic synonyms in specific terms – i.e., *Absent middle phalanges* listed as synonym of *Aplasia of the middle phalanges of the hand*.

2. The logical definition of certain PFO Anatomical Composites introduces noise. For example, a simplified logical definition of `Middle phalanges`, without

considering cardinality and anatomical coordinate, is the union of `FMA: Phalanx of finger` and `FMA: Phalanx of toe`. The mapping process uses both underpinning FMA concepts for lexicon generation and hence creates an entire series of false positives. A slightly different example is `Pubic rami`, which is an union of `FMA: Inferior pubic ramus` and `FMA: Superior pubic ramus`.

3. Some of the FMA synonyms are too verbose and thus lead to false positives. For example, `FMA: Diaphysis` has listed *Body of long bone* as synonym. While *Diaphysis* is a fairly unique term and achieves a low similarity score against almost every other anatomical part (based on our function), our optimal trace calculation method may choose its synonym as a better pick, due to the high similarity values it produces (even if it is always penalised).

The broader mapping has been only partly affected by the above issues, since some of them are discarded when looking for the lowest common ancestor, and has achieved a maximal P@1 of 0.91 for the simple and 0.72 for the composite anatomy category. However, here the issues have shifted towards the second challenge mentioned in Section 2, i.e., the structure of HPO. Firstly, the multiple inheritance aspect of HPO influences the computation of the lowest common ancestor (LCA). Consequently, this may lead to the pruning of an entire set of relevant HPO terms in the advantage of a LCA which represented by a set of false positives. Subsequently, this is used in the final similarity computation where it achieves a very low score (see example in Section 2). Secondly, the distribution of certain phenotypes is heavily skewed and produces an almost linear branch of the corresponding abnormality. As a result, the algorithm picks concepts that are more specific than required. For example, `Retarded_ossification_of_the_femoral_neck` will be associated with the broader concept `HP_0006429` (*Broad femoral neck*) instead of `HP_0003367` (*Abnormality of the femoral neck* – the super class of `HP_0006429`), because it has a larger number of relevant children than relevant siblings.

Unfortunately, most of the issues listed above cannot be cleanly fixed from an algorithmic perspective without introducing specific work-arounds. On the HPO side, we will work together with the HPO maintainers on addressing the aspects we have mentioned. On our side, we will focus on making the algorithm more robust, in order to deal especially with the verbosity of the FMA synonyms.

6 Conclusion

In this paper we have reported on our efforts in creating a standard representation for radiographic findings in the skeletal dysplasia domain and mapping concepts modeled in this representation to Human Phenotype Ontology terms . We have shown that our Phenotype Fragment Ontology provides a flexible metamodel that bridges the diverse lexical groundings of the radiographic findings by using widely adopted ontologies to underpin the actual concepts definition. Subsequently, we have described an exact and a broader matching algorithm able to efficiently map PFO concepts to HPO terms. From an application perspective, the SKELETOME platform now uses as part of its semantic annotation

process both HPO terms, and PFO instances, in particular where appropriate HPO terms are not available. Future work will focus on extending PFO to cater for a series of particular cases, such as combinations of atomic phenotypes, e.g., *Acromesomelic brachymelia*.

Acknowledgments. The work presented in this paper is supported by the Australian Research Council (ARC) under the Discovery Early Career Researcher Award (DECRA) – DE120100508 and the Linkage grant SKELETOME – LP100100156. The experiments detailed in the paper have used material from "*Bone dysplasia: an atlas of genetic disorders of skeletal development*" by Jurgen W. Spranger, Paula W. Brill, Andrew K. Poznanski (2002) – By permission of Oxford University Press, Inc.

References

1. Groza, T., Zankl, A., Li, Y.-F., Hunter, J.: Using Semantic Web Technologies to Build a Community-Driven Knowledge Curation Platform for the Skeletal Dysplasia Domain. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part II. LNCS, vol. 7032, pp. 81–96. Springer, Heidelberg (2011)
2. Robinson, P.N., Kohler, S., Bauer, S., Seelow, D., Horn, D., Mundlos, S.: The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. The American Journal of Human Genetics 83(5), 610–615 (2008)
3. Rosse, C., Mejino, J.L.V.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J. of Biomed. Inf. 36(6), 478–500 (2003)
4. Gkoutos, G.V., et al.: Entity/Quality-Based Logical Definitions for the Human Skeletal Phenome using PATO. In: Proceedings of the 31st Annual International Conference of the IEEE EMBS, Minneapolis, Minnesota, USA, pp. 7069–7072 (2009)
5. Hu, W., Qu, Y., Cheng, G.: Matching large ontologies: A divide-and-conquer approach. IEEE Data and Knowledge Engineering 67(1), 140–160 (2008)
6. Li, J., et al.: RiMOM: A dynamic multi-strategy ontology alignment framework. IEEE Data and Knowledge Engineering 21(8), 1218–1232 (2009)
7. Lambrix, P., Tan, H.: SAMBO – A system for aligning and merging biomedical ontologies. Journal of Web Semantics 4(1), 196–206 (2006)
8. Nagy, M., Vargas-Vera, M.: Multi-Agent Ontology Mapping Framework for the Semantic Web. IEEE Transactions on Systems, Man, and Cybernetics – Part A 41(4), 693–704 (2011)
9. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering (2012)
10. Jonquet, C., et al.: The open biomedical annotator. In: Proc. of the 2010 AMIA Summit of Translational Bioinformatics, San Francisco, CA, US, pp. 56–60 (2010)