A Preliminary Ontology for Spermatozoa Analysis

Leonardo A. Molas[1], Natalia V. Revollo Sarmiento[1], Sergio Alejandro Gómez[2], Rolando González-José[3], and Claudio A. Delrieux[1]

[1] Laboratorio de Ciencias de las Imágenes, Universidad Nacional del Sur (UNS)
Av. Alem 1253, Bahía Blanca, Buenos Aires, Argentina
[2] Instituto de Ciencias e Ingeniería de la Computación (ICIC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Av. Alem 1253, Bahía Blanca, Buenos Aires, Argentina
[3] Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Bvd. Brown 2915. U9120ACD, Puerto Madryn, Argentina

Abstract. Biomedical computation has used ontologies as a foundation for building knowledge-based systems and technologies for Semantic Web applications. However, so far they had not been utilized in andrology related domains. This paper presents a preliminary effort to provide a comprehensive ontology for classifying and managing spermatozoa samples and their phenotypic traits, in order to analyze and diagnose them. Our study is centered around the development of a Computer Asisted Sperm Analysis (CASA) system.

Keywords: Ontologies, biomedical applications, andrology, spermatozoa classification

1 Introduction

The evaluation of the spermatozoa's imagery, also known as spermograms, is a fundamental study in andrologic clinical analysis that allows to study the male fertility and related pathologies [21]. The exam is achieved by counting the amount of spermatozoa that is considered to be healthy according to the standards of the World Health Organization (WHO) [22] within microscopic images of sperm samples. Given the variability underlying microscopy imaging, and the complexity of the shape and form of the spermatozoa, this study must be performed by specialized andrologists [7, 17], and therefore it is time-consuming, prone to human subjective variability and error, and impractical since most clinical analysis laboratories do not usually count with such specialists. Consequently, several studies have been conducted to automatize this process, given its procedural nature, using dedicated hardware and software in what it is known as Computer Assisted Sperm Analysis (CASA) [18, 9].

However, these studies have been done independently from other biomedical computation endeavours [23]. Significant research have been conducted in the

pursuit of a homologation of medical terminologies (or vocabularies), such as the Unified Medical Language System (UMLS) [3], the Gene Ontology (GO) [6] and the Open Biomedical Ontology (OBO Foundry) [24], in order to facilitate its usage through computational means [8]. In particular, biomedical informatics is based on using both ontologies and ontology design [20], which allow for natural language processing tasks such as automatic summarization [5], text annotation [4], and word sense disambiguation [19], among others.

In this work we propose a preliminary study for sperm analysis using a comprehensive and well-defined ontology. This stuy is part of an ongoing project developed by a group of computer scientists in conjunction with biologists and andrologists to elaborate a complete system for automatic sperm analysis and diagnosis, based on image processing of human sperm samples. The project is aimed towards the detection of male human infertility, but also may be applicable in the evaluation of a vasectomy procedure, as well as an early detection of many andrology-related pathologies such as infection in the male reproductive system or testicular cancer.

The rest of this paper is structured as follows. In Section 2, we briefly recall the main concepts of knowledge representation with Description Logics. Section 3 presents a preliminary version of a Description Logic ontology for modeling the characteristics of both normal and anormal human spermatozoa. Section 4 includes a discussion of the the scope and limitations of our approach. Section 5 discusses related work. Finally, in Section 6, we highlight the conclusions and propose future research perspectives.

2 Description Logics

Description Logics (DLs) are a family of knowledge representation languages that can be used to represent knowledge of an application domain in a structured and formally well-understood way [1]. The important notions of the domain are described by concept descriptions, *i.e.*, expressions that are built from atomic concepts (unary predicates) and atomic roles (binary predicates) using the concept and role constructors provided by the particular DL.

Concept descriptions can be used to build statements in a DL knowledge base, which typically come in two parts, a terminological, or TBox, where we can describe the relevant notions of an application domain by stating properties of concepts and roles; and an assertional one, or ABox, that is used to describe a concrete situation by stating class membership of individuals as well as properties of individuals.

This paper is centered around the level of expressiveness of $\mathcal{SHOIN}(\mathcal{D})$, on which the ontological language OWL-DL is based [13], which is similar to the well known $\mathcal{SHOQ}(\mathcal{D})$ DL [14], but is extended with inverse roles ($\mathcal{I}$) and restricted to unqualified number restrictions ($\mathcal{N}$).

3 A Comprehensive Ontology for the Viable Sperm Problem

The criteria for determining whether a human sperm sample is classified as normal or abnormal are well specified in [22]. These criteria are described in natural language and for this reason they are clear to a human specialist who performs the analysis of sperm samples. To automatize the process of classification, we need to make these criteria machine-processable. We present here our preliminary efforts in developing an ontology for modeling them.

The observations on spermatozoa recovered from the female reproductive tract have helped to define the appearance of potentially fertilizing (*i.e.*, morphologically normal) spermatozoa, despite the variable morphology that makes assessment difficult. By the strict application of certain criteria of sperm morphology, relationships between the percentage of normal forms and various fertility endpoints have been established, as detailed in [22]. Particularly in [22, Section 2.15.1], a description of the principles for abnormal sperm classification is laid out, in a way that it is plausible to formulate an ontology that models it.

A comprehensive set of concepts and axioms to devise a representation for spermatozoa morphology and its normal characteristics leads to an ontology of a substancial size (which prevents from laying it out entirely in this paper). Nonetheless, the most relevant parts will be used throughout this work to illustrate the most salient concepts (a complete version of the ontology can be found in `https://github.com/leosmolas/sperm-ontology`).

3.1 Structure of the Knowledge Base

The features required to classify normal/abnormal spermatozoa with respect to its morphology are represented in the knowledege base in a structured form. The concepts and their relationships are laid out in Fig. 1.

In the diagram, each entity maps to a concept in the ontology, and each relationship to a relationship in there as well. Each concept has their respective attributes, e.g. $MidPiece$ has an attribute $midpieceLength$ that stores the mid-piece length. The diagram is symmetric, since it lower half is a mirrored version of the upper, and each of the concepts has an *is-a* relationship with its correspondent. Only one of this relationships is shown ($Spermatozoon$ / $NormalSpermatozoon$) in order to simplify the diagram.

The upper-half will contain the data of all the spermatozoa and its parts, and the lower, only the normal ones. To perform this classification, the ontology has also axioms that represent relations between each concept and its normal version considering their attributes. This will be explained in more detail below.

Each concept, relationship, attribute and datatype is listed in the alphabet Σ. Concepts and datatypes are in $UpperCamelcase$ (i.e. $Spermatozoon$, $PrincipalPiece$, $Integer$), while both relationships and attributes are in regular $camelCase$ (e.g. $spermHead$, $headSize$). Relationships between two concepts are named by the concatenation of their names, and may be shortened (e.g., the

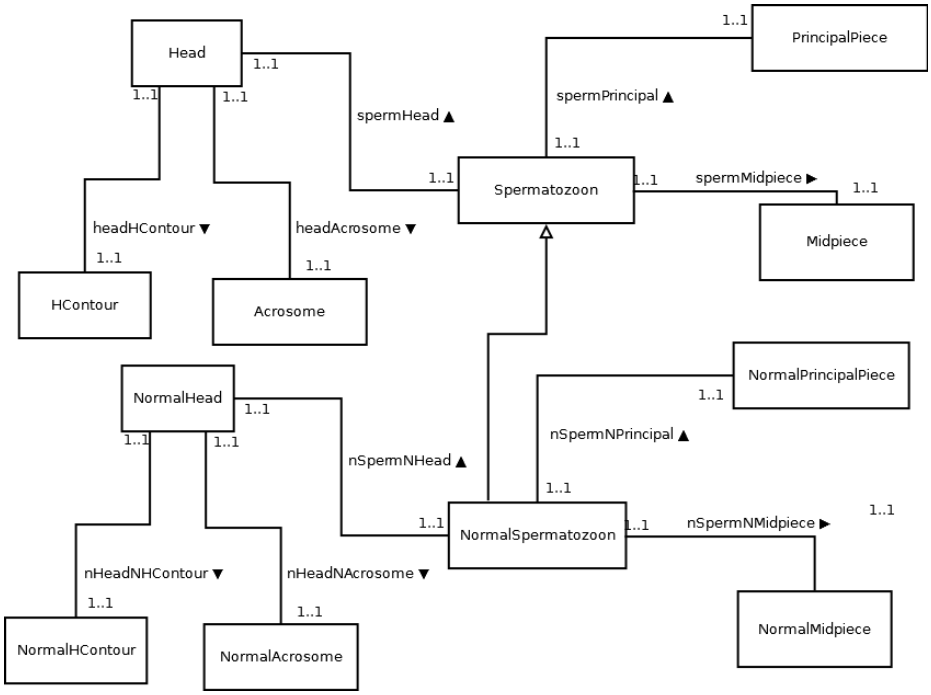

Fig. 1. Entity-Relationship diagram that models the structure of the knowledge base

$$
\begin{aligned}
Spermatozoon &\sqsubseteq \exists spermPrincipal \sqcap (\geq 1\, spermPrincipal)\\
\exists spermPrincipal &\sqsubseteq Spermatozoon\\
PrincipalPiece &\sqsubseteq \exists spermPrincipal^{-} \sqcap (\geq 1\, spermPrincipal^{-})\\
\exists spermPrincipal^{-} &\sqsubseteq PrincipalPiece\\
Spermatozoon &\sqsubseteq \exists spermMidpiece \sqcap (\geq 1\, spermPrincipal)\\
\exists spermMidpiece &\sqsubseteq Spermatozoon\\
Midpiece &\sqsubseteq \exists spermMidpiece^{-} \sqcap (\geq 1\, spermPrincipal^{-})\\
NormalSpermatozoon &\sqsubseteq Spermatozoon
\end{aligned}
$$

Fig. 2. Extract of the section of the ontology that models the shape of the knowledge base.

$$
\begin{aligned}
\exists principalUniform &\sqsubseteq PrincipalPiece\\
\exists principalUniform^{-} &\sqsubseteq Boolean\\
\exists principalWidth &\sqsubseteq PrincipalPiece\\
\exists principalWidth^{-} &\sqsubseteq Float\\
\exists principalSharpAngle &\sqsubseteq PrincipalPiece\\
\exists principalSharpAngle^{-} &\sqsubseteq Boolean\\
\exists principalLoop &\sqsubseteq PrincipalPiece\\
\exists principalLoop^{-} &\sqsubseteq Boolean
\end{aligned}
$$

Fig. 3. Extract of the section of the ontology that lists the attributes of the concepts with their datatypes.

relationship between $Spermatozoon$ and $Midpiece$ is noted as $spermMidpiece$). The normal counterpart of each concept has the $Normal-$ prefix (e.g. $Head$ and $NormalHead$), while the relationships between them shortens from $Normal-$ to $n-$ and $-N-$ (e.g. the relationship between $NormalSpermatozoon$ and $NormalMidpiece$ is noted as $nSpermNMidpiece$).

The first part of the ontology establishes the shape of the knowledge base, and merely describes the diagram in Fig. 1, considering concepts, relationships and multiplicity constraints. An excerpt of it can be found in Fig. 2. All the attributes are also listed in the ontology, and were selected considering the guidelines described in [22]. We present several examples in Fig. 3. Their importance will be explained below (see Section 3.2).

3.2 Classification Axioms

Thus far, the ontology has been filled with axioms that model the entity-relationship diagram in Figure 1. This section will focus on the axioms that model

the criteria applied by andrologists for classificating spermatozoids. Specifically, quotes from [22] [4] will be followed by excerpts from the ontology which models them.

> *The head should be smooth, regularly contoured and generally oval in shape.*

$$\begin{aligned} &\exists hcontourSmooth.\{true\} \sqcap \\ &\exists hcontourOvalShape.\{true\} \sqcap \\ &\exists hcontourRegular.\{true\} \equiv NormalHContour \end{aligned}$$

The strategy for mirroring the requirements is, for each concept referred to by them, selecting the attributes of it that are considered, and limiting the range of the correspondent datatype to the specified boundaries. In the above example, the characteristics from the quote are modeled by boolean attributes that are contained in $HContour$ (i.e. $\exists hcontourSmooth^{-} \sqsubseteq HContour$). They are required to be true to be considered normal, so the range $\{true\}$ is selected for each of them, and finally their conjunction is generated to conform the concept $NormalHContour$.

> *There should be a well-defined acrosomal region comprising 40–70% of the head area (. . .). The acrosomal region should contain no large vacuoles, and not more than two small vacuoles, which should not occupy more than 20% of the sperm head.*

$$\begin{aligned} &\exists acrosomeCoverage. \leq 40 \sqcap \\ &\exists acrosomeCoverage. \geq 70 \sqcap \\ &\exists acrosomeLargeVacuoles.\{false\} \sqcap \\ &\exists acrosomeVacuolesQuantity. \leq 2 \sqcap \\ &\exists acrosmeVacuoleCoverage. \leq 20 \ \equiv NormalAcrosome \end{aligned}$$

In this example, $acrosomeCoverage$ and $acrosomeVacuolesQuantity$ are attributes of numeric datatypes, and their boundaries are set with *atleast* and *atmost* operators. The following examples conclude this part of the ontology.

> *The post-acrosomal region should not contain any vacuoles.*

[4] Reprinted from Laboratory manual for the Examination and processing of human semen, Fifth Edition, World Health Organization, Department of Reproductive Health and Research, Chapter 2 “Standard Procedures”, pages 68 and 69, 2010.

The head dimensions of 77 Papanicolaou-stained spermatozoa (...) , measured by a computerized system (coefficient of variation for repeated measurements 2–7%) had the following dimensions: median length 4.1m, 95% CI 3.7–4.7; median width 2.8 µm, 95% CI 2.5–3.2; median length-to-width ratio 1.5, 95% CI 1.3–1.8.

$$\begin{aligned} \exists headVacuolesQuantity.\{0\} \sqcap \\ \exists headLength. \geq 3.7 \sqcap \\ \exists headLength. \leq 4.7 \sqcap \\ \exists headWidth. \geq 2.5 \sqcap \\ \exists headWidth. \leq 3.2 \sqcap \\ \exists headAspectRatio. \geq 1.3 \sqcap \\ \exists headAspectRatio. \leq 1.8 \equiv NormalHead \end{aligned}$$

The midpiece should be slender, regular and about the same length as the sperm head.

The midpieces of 74 Papanicolaou-stained spermatozoa (...) and measured by the same computerized system had the following dimensions: median length 4.0µm, 95% CI 3.3–5.2; median width 0.6µm, 95% CI 0.5–0.7.

$$\begin{aligned} \exists midpieceRegular.\{true\} \sqcap \\ \exists midpieceSlender.\{true\} \sqcap \\ \exists midpieceLenght. \geq 3.3 \sqcap \\ \exists midpieceLenght. \leq 5.2 \sqcap \\ \exists midpieceWidth. \geq 0.5 \sqcap \\ \exists midpieceWidth. \leq 0.7 \equiv NormalMidpiece \end{aligned}$$

The principal piece should have a uniform calibre along its length, be thinner than the midpiece and be approximately 45µm long (about 10 times the head length) It may be looped back on itself (...), provided there is no sharp angle indicative of a flagellar break.

$$\begin{aligned} \exists principalUniform.\{true\} \sqcap \\ \exists principalSharpAngle.\{false\} \equiv NormalPrincipalPiece \end{aligned}$$

4 Discussion

The ontology presented was originally conceived for its utilization within the scope of a CASA system that we are currently developing. Nonetheless, it would be feasible to consider this ontology as an independent module to be used in other applications, such as natural language processing applications and Semantic Web applications. Currently, the ontology presented in this paper is still under development, despite the fact that we have already covered the most relevant sections of the authoritative reference [22]. This partial completeness is due to different criteria among biologists, andrologists and genetists. In other words, criteria from different scientific endeavours provide varying standards regarding classification of healthy or unhealthy sperms. Moreover, these criteria may vary over time, since further biological research or analysis may provide evidence for changing or enhancing priorly adopted standards. For this reason, the ontology presented here may provide means for reconciling these different or varying standards, in a evolutionary multi-agent system that may suggest different classifications that follow these diverse criteria. Possible alternatives to achieve this goal are under consideration, and discussed in Section 6.

5 Related Work

Notwithstanding the importance of the problem, to the best of our knowledge, there has not been another study utilizing ontologies for CASA systems. Nonetheless, several researches had been conducted utilizing ontologies for bioinformatics.

As mentioned, OBO Foundry is a collaborative endeavour that mantains a suite of orthogonal interoperable reference ontologies in the biomedical domain [24]. Köhler, Doelken, Mungall et al. presents in [15] the Human Phenotype Ontology (HPO), which provides a structured, comprehensive and well-defined ontology describing human phenotypic abnormalities. They have developed logical definitions for almost half of all HPO classes using terms from other OBO Foundry ontologies, allowing interoperability between them.

Köhler, Doelken, Ruef et al. generated a cross-species phenotype ontology for human, mouse and zebrafish that contains classses from the HPO and the Mammalian Phenotype Ontology in [16]. They also provide annotation data connecting human genes to phenotype classes from the ontology, with the respective data generation pipeline.

Finally, Bandrowski et al. in [2] presents the Ontology for Biomedical Investigations (OBI), which provides terms with precisely defined meanings to describe all aspects of how investigations in the biological and medical domains are conducted. It reuses ontologies from the OBO Project and adds the ability to describe how that knowledge was derived. OBI allows the use of a single internally consistent resource that can be applied to multiple types of experiments for this type of applications, by creating terms for entities involved in biological

and medical investigations and by importing parts of other biomedical ontologies such as the Gene Ontology (GO), Chemical Entities of Biological Interest (ChEBI) and Phenotype Attribute and Trait Ontology (PATO).

6 Conclusions and Future Work

We have presented a preliminary ontology for storing the phenotypic traits of spermatozoa and for its posterior normal/abnormal classification. As discussed before, one of the difficulties in this study is given by the consideration of the different criteria of the experts involved. A possible solution is to consider this ontology as modular, i.e. the representation of the entity-relationship diagram as the central module, keeping the knowledge base and the classification axioms as separered modules that could be interchangeable. This would allow us to create several modules with different classification criteria that would possibly generate different results and diagnostics.

Another approach would be the utilization of different kinds of description logics with the capabilities of storing paraconsistent or inconsistent data. This would open the spectrum to the utilization of non-monotonic reasoning, considering the fact that the criteria and results of diagnostics change among different expert biologists. In this regard, previous work of Gómez et al. [10, 12, 11] could be useful for these alternative avenues of research.

The possibility diagnosing pathologies other than infertility might be possible considering and analysing the different types of malformations present in the sperm morphology. An extensive list is presented in [22, Table 2.6], with explanations of the different malformations, though it might not be complete. Moreover, [22, Section 2.15.2] provides a classification for abnormal sperm morphology; these categories of defects could be used to help in the diagnostics of pathologies, and can be reflected through new axioms in the ontology.

Finally, the collection of ontologies OBO Foundry could be a home for this study. For this, it would be necessary to translate our work to a different type of language, and also to do modifications to it to satisfy their standards and principles in order to assure integration and interoperability. This would allow our ontology to join a well-known initiative and our efforts to rely on a strong community of developers.

Acknowledgements. The work of Sergio Gómez and of Claudio Delrieux are partially funded by Secretaría General de Ciencia y Técnica, Universidad Nacional del Sur, Argentina (grants PGI: 24/N040 and PGI: 24/K061).

References

[1] Baader, F., Horrocks, I., Sattler, U.: Description logics. In: Handbook on ontologies, pp. 21–43. Springer (2009)

[2] Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., Chibucos, M.C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., et al.: The ontology for biomedical investigations. PloS one 11(4), e0154556 (2016)

[3] Bodenreider, O.: The unified medical language system (umls): integrating biomedical terminology. Nucleic acids research 32(suppl 1), D267–D270 (2004)

[4] Carr, L., Hall, W., Bechhofer, S., Goble, C.: Conceptual linking: ontology-based open hypermedia. In: Proceedings of the 10th international conference on World Wide Web. pp. 334–342. ACM (2001)

[5] Chiang, J.H., Shin, J.W., Liu, H.H., Chin, C.L.: Genelibrarian: an effective gene-information summarization and visualization system. BMC bioinformatics 7(1), 1 (2006)

[6] Consortium, G.O., et al.: The gene ontology (go) database and informatics resource. Nucleic acids research 32(suppl 1), D258–D261 (2004)

[7] Cooper, T.G., Noonan, E., Von Eckardstein, S., Auger, J., Baker, H.G., Behre, H.M., Haugen, T.B., Kruger, T., Wang, C., Mbizvo, M.T., et al.: World health organization reference values for human semen characteristics. Human reproduction update p. dmp048 (2009)

[8] Díaz-Galiano, M.C., Martín-Valdivia, M.T., Ureña-López, L.: Query expansion with a medical ontology to improve a multimodal information retrieval system. Computers in biology and medicine 39(4), 396–403 (2009)

[9] Farrell, P., Presicce, G., Brockett, C., Foote, R.: Quantification of bull sperm characteristics measured by computer-assisted sperm analysis (casa) and the relationship to fertility. Theriogenology 49(4), 871 – 879 (1998), http://www.sciencedirect.com/science/article/pii/S0093691X98000363

[10] Gómez, S.A., Chesñevar, C.I., Simari, G.R.: Reasoning with Inconsistent Ontologies Through Argumentation. Applied Artificial Intelligence 1(24), 102–148 (2010)

[11] Gómez, S.A., Chesñevar, C.I., Simari, G.R.: ONTOarg: A Decision Support Framework for Ontology Integration based on Argumentation. Expert Systems with Applications 40, 1858–1870 (2013)

[12] Gómez, S.A., Simari, G.R.: Merging of ontologies using belief revision and defeasible logic programming. Inteligencia Artificial 16(52), 16–28 (2013)

[13] Horrocks, I.: Owl: A description logic based ontology language. In: Logic Programming, pp. 1–4. Springer (2005)

[14] Horrocks, I., Sattler, U.: Ontology reasoning in the shoq (d) description logic. In: IJCAI. vol. 1, pp. 199–204 (2001)

[15] Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C., Brown, D.L., Brudno, M., Campbell, J., et al.: The

human phenotype ontology project: linking molecular biology and disease through phenotype data. Nucleic acids research 42(D1), D966–D974 (2014)
[16] Köhler, S., Doelken, S.C., Ruef, B.J., Bauer, S., Washington, N., Westerfield, M., Gkoutos, G., Schofield, P., Smedley, D., Lewis, S.E., et al.: Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research 2 (2014)
[17] Maree, L., Du Plessis, S., Menkveld, R., Van der Horst, G.: Morphometric dimensions of the human sperm head depend on the staining method used. Human Reproduction 25(6), 1369–1382 (2010)
[18] Marshburn, P., McIntire, D., Carr, B., Byrd, W.: Spermatozoal characteristics from fresh and frozen donor semen and their correlation with fertility outcome after intrauterine insemination. Fertility and sterility 58(1), 179–186 (1992)
[19] Martín-Valdivia, M.T., Ureña-López, L., García-Vega, M.: The learning vector quantization algorithm applied to automatic text classification tasks. Neural Networks 20(6), 748–756 (2007)
[20] Musen, M.A., et al.: Medical informatics: searching for underlying components. Methods of information in medicine 41(1), 12–19 (2002)
[21] Ombelet, W., Menkveld, R., Kruger, T., Steeno, O.: Sperm morphology assessment: historical review in relation to fertility. Human Reproduction Update 1(6), 543–557 (1995)
[22] Organization, W.H.: WHO laboratory manual for the examination and processing of human semen. World Health Organization (2010)
[23] Shortliffe, E.H., Cimino, J.J.: Biomedical informatics: computer applications in health care and biomedicine. Springer Science & Business Media (2013)
[24] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., et al.: The obo foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology 25(11), 1251–1255 (2007)