HAL open science

Sustainable Formal Representation Of Breast Cancer Grading Histopathological Knowledge

K. Traore, C. Daniel, M.-C. Jaulent, T. Schrader, Daniel Racoceanu, Y. Kergosien

To cite this version:

K. Traore, C. Daniel, M.-C. Jaulent, T. Schrader, Daniel Racoceanu, et al.. Sustainable Formal Representation Of Breast Cancer Grading Histopathological Knowledge. Diagnostic Pathology, 2016, 9 (1), 10.17629/www.diagnosticpathology.eu-2016-8:154 . hal-01366742

HAL Id: hal-01366742

https://hal.sorbonne-universite.fr/hal-01366742v1

Submitted on 15 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

the diagnostic pathology journal DIAGNOSTIC PATHOLOGY

13th European Congress on Digital Pathology Proceedings, diagnostic pathology 2016,8:154
ISSN 2364-4893
DOI: http://dx.doi.org/10.17629/www.diagnosticpathology.eu-2016-8:154

Proceedings

SY05.02 | **Computer Aided Diagnosis**

Sustainable Formal Representation Of Breast Cancer Grading Histopathological Knowledge

K. Traore*[1, 2, 3], C. Daniel[2, 4], M.-C. Jaulent[2], T. Schrader[5], D. Racoceanu[3], Y. Kergosien[2, 6]

[1]Université Pierre Marie Curie, UPMC-Paris 6, LIMICS: INSERM U 1142, LIB: CNRS UMR 7371, INSERM U 1146, Paris, France, [2]Sorbonne Universités, UPMC Univ Paris 06, INSERM, Université Paris 13, Sorbonne Paris Cité, Laboratoire d'Informatique Médicale et Ingénierie des Connaissances en eSanté (LIMICS - UMR_S 1142), 15 rue de l'école de médecine, Paris, France, [3]Sorbonne Universités, UPMC Univ Paris 06, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale (LIB), 75013, Paris, France, [4]Assistance Publique-Hôpitaux de Paris (AP-HP), CCS SI Patient, Paris, France, [5]University of Applied Sciences Brandenburg Magdeburger, Department Informatics and Media, Brandenburg, Germany, [6]Département d'Informatique Université de Cergy-Pontoise, Cergy-Pontoise, France

Introduction/ Background

Recently, histopathology has seen the introduction of several tools such as slide scanners and virtual slide technologies, creating the conditions for broader adoption of computer aided diagnosis based on whole slide images (WSI) to reduce observation variability between pathologists. This change brings up a number of new scientific challenges such as the sustainable management of the semantics associated to the grading process, image analysis and annotation in order to facilitate pre-filled report generation. The College of American Pathologists cancer checklists and protocols (CAP-CC&P) [1] are reference resources for complete Anatomic Pathology (AP) reporting of malignant tumors. Current terminology systems for AP structured reporting gather terms of very different granularity [2, 3] and have not yet been compiled in a systematic approach. Semantic data models are formal representations of knowledge in a given domain that allow both human users and software applications to consistently and accurately interpret domain terminology [4, 5].

Aims

Our objective is to i) analyze the histopathological knowledge for breast cancer grading available in the reference CAP CC&P and ii) to build a sustainable formal representation of this knowledge based on existing biomedical ontologies in NCBO Bioportal [6, 7] and UMLS semantic types [8].

Methods

Our methodology was first experimented in the context of two cancer grading methods for invasive (Nottingham system) and ductal in situ breast carcinoma. A corpus consisting of 5 texts or "notes" was first selected by an AP expert from the two corresponding CAP CC&Ps. From each note the expert also extracted a list of key concepts to be used as a "gold standard". We used NCBO Annotator [9] for automatic analysis of the corpus. Annotator supports the biomedical community in tagging raw texts automatically with concepts from relevant biomedical ontology and terminology repositories. The methodology used consists in:

i) Automatic textual analysis and annotation of the corpus based on the 417 ontologies available on the NCBO platform. We selected a subset of ontologies based on the number of identified concepts and evaluated their relevancy with respect to the gold standard.

ii) Semantic modeling of the automatically extracted concepts into a sustainable formal representation based on their UMLS semantic types.

the diagnostic pathology journal DIAGNOSTIC PATHOLOGY

13th European Congress on Digital Pathology Proceedings, diagnostic pathology 2016,8:154
ISSN 2364-4893
DOI: http://dx.doi.org/10.17629/www.diagnosticpathology.eu-2016-8:154

Results

We identified NCIT, SNOMED-CT, NCI CaDSR Values set, LOINC and PathLex as the ontologies providing the highest number of annotated concepts. *Table 1* shows as percentages the coverages of the concepts of each note by the annotations of the 5 reference ontologies. Percentages can add to more than 100 for a single note due to the possible overlap in ontologies coverages. *Table 2* uses the same format when only concepts from the gold standards are counted to quantify annotations. From the list of extracted concepts, we made a preliminary formal representation of the histopathological knowledge based on the UMLS semantic types of concepts. *Figure 1* shows the so proposed semantic modeling in the context of tubular differentiation. The novelty of this approach is the federation of the knowledge issued from different sources (CAP CC&P, NCBO ontologies and UMLS Metathesaurus) and the sustainable management of the associated semantics. This opens the perspective of building an AP observation ontology that will allow an accurate representation of AP reports understandable by both human and software applications.

		Ontologies									
		NCIT		SNOMED-CT		NCI caDSR Value Sets		LOINC		PathLex	
	Number of Concepts	Concepts	Coverage (%)	Concepts	Coverage (%)	Concepts	Coverage (%)	Concepts	Coverage (%)	Concepts	Coverage (%)
Note#1	**23**	**15**	65	**10**	43	**5**	22	**6**	26	**1**	4
Note#2	**102**	**68**	67	**40**	39	**25**	25	**18**	18	**3**	3
Note#3	**58**	**31**	53	**28**	48	**15**	26	**12**	21	**6**	10
Note#4	**47**	**33**	70	**17**	36	**6**	13	**11**	23	**3**	6
Note#5	**103**	**71**	69	**31**	30	**29**	28	**12**	12	**3**	3
	Total	**Total**	Average coverage (%)	**Total**	Average coverage (%)	**Total**	Average coverage (%)	**Total**	Average coverage (%)	**Total**	Average coverage (%)
	333	**218**	65	**126**	38	**80**	24	**59**	18	**16**	5

Table 1.

Gold standard		**Ontologies**									
		NCIT		LOINC		NCI caDSR Value Sets		SNOMED-CT		PathLex	
	Number of Concepts	Concepts	Coverage (%)	Concepts	Coverage (%)	Concepts	Coverage (%)	Concepts	Coverage (%)	Concepts	Coverage (%)
Note#1	**2**	**2**	100	**2**	100	**0**	0	**0**	0	**0**	0
Note#2	**8**	**4**	50	**0**	0	**0**	0	**0**	0	**0**	0
Note#3	**14**	**7**	50	**1**	7	**0**	0	**1**	7	**1**	7
Note#4	**5**	**4**	80	**3**	60	**1**	20	**1**	20	**1**	20
Note#5	**26**	**10**	38	**1**	4	**6**	23	**4**	15	**1**	4
	Total	**Total**	Average coverage (%)	**Total**	Average coverage (%)	**Total**	Average coverage (%)	**Total**	Average coverage (%)	**Total**	Average coverage (%)
	55	**27**	49	**7**	13	**7**	13	**6**	11	**3**	5

Table 2.

the diagnostic pathology journal DIAGNOSTIC PATHOLOGY

13th European Congress on Digital Pathology Proceedings, diagnostic pathology 2016,8:154
ISSN 2364-4893
DOI: http://dx.doi.org/10.17629/www.diagnosticpathology.eu-2016-8:154

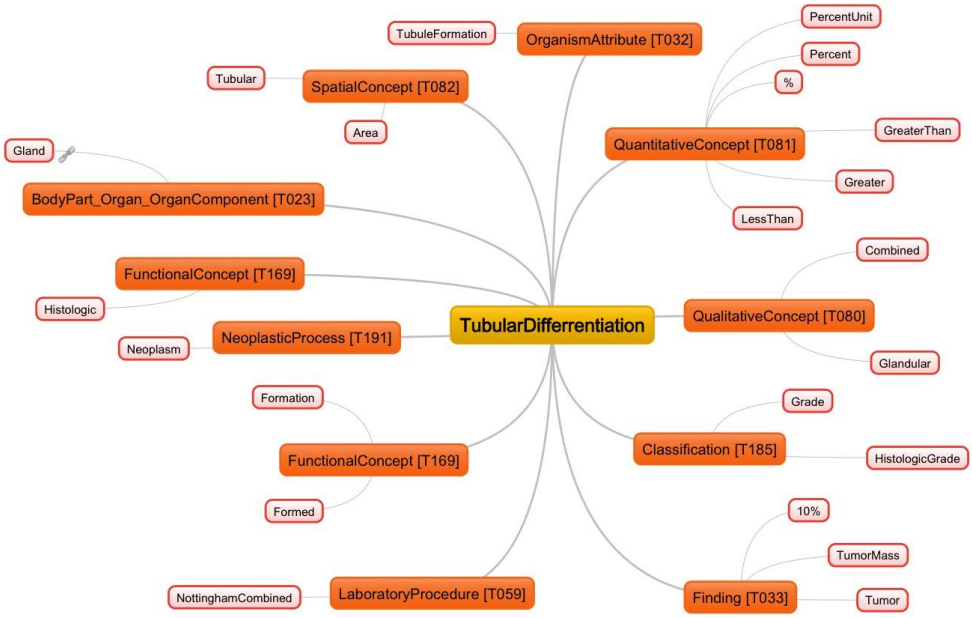

Figure 1.

References:

[1] College of American Pathologists, Cancer Protocols and Checklists, 2013: DCIS – Breast Revised: December 18, 2013 Version: 3.2.0.0, Invasive Breast Posted: December 18, 2013 Version:3.2.0.0 , available from: http://www.cap.org/

[2] Daniel C., Booker D., Beckwith B., Della Mea V., García-Rojo M., Havener L., Kennedy M., Klossa J., Laurinavicius A., Macary F., Punys V., Scharber W., Schrader T., Standards and specifications in pathology: image management, report management and terminology, Stud Health Technol Inf. 2012, 179: 105–122.

[3] Haroske G., Schrader T. , A reference model based interface terminology for generic observations in Anatomic Pathology Structured Reports, Diagnostic Pathology 2014, 9 (1):S4.

[4] Bodenreider O., Biomedical ontologies in action: role in knowledge management, data integration and decision support, Yearb Med Inform. 2008:67-79.

[5] Rubin D.L., Shah N.H., Noy N.F., Biomedical ontologies: a functional perspective, Briefings in Bioinformatics 2008, 9(1): 75-90.

[6] Musen M.A., Noy N.F., Shah N.H., Whetzel P.L., Chute C.G., Story M.A., Smith B.; NCBO team, The National Center for Biomedical Ontology, J Am Med Inform Assoc. 2012, 19(2):190-5. Epub 2011 Nov 10.

[7] Whetzel P.L., Noy N.F., Shah N.H., Alexander P.R., Nyulas C., Tudorache T., Musen M.A., Bioportal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, Nucleic Acids Res. 2011, 39(Web Server issue):W541-5. Epub 2011 Jun 14

[8] Bodenreider O., The Unified Medical Language System (UMLS): integrating biomedical terminology [Internet], [Accessed: 17-Dec-2015, Avalable from: http:/nar.oxfordjournals.org

[9] Jonquet C., Shah N.H., Musen M.A., The open biomedical annotator, Summit on Translat Bioinforma. 2009, 1:56-60.