Discovering Semantic Relationships between NCD and Lifestyle Patterns using Ontologies

María J. Somodevilla, Ismael Mena, Ivo H. Pineda, Concepcion Perez de Celis

Computer Science Department, Puebla,
Mexico

{mariajsomodevilla, imenav85, ivopinedatorres, mcpcelish,
dvilarinoayala}@gmail.com

Abstract. The volume of biomedical spatial information available on line increases day by day, which need to be exploited and shared by users in different knowledge areas. NCD[1] kill 38 million people each year and considering the seriousness of the problem globally, there are already underway government strategies to reduce risk factors and early detection and timely treatment. In this article, it's discussed the problem of the NCD using Semantic Web tools. The proposed system uses six ontologies which formalize concepts related to people, physical activity, the NCD, nutrition, geographic regions and symptoms to give information about lifestyle patterns. SWRL[2] rules are used to define accurate axioms which allow improving classification of the individuals.

Keywords: Ontologies system, SWRL rules, lifestyle patterns, NCD.

1 Introduction

NCDs kill 38 million people each year and are classified as cardiovascular, cancer, respiratory and diabetes. Affect all age groups and all geographic regions. The risk factors associated with NCDs are unhealthy stuff consumption, physical inactivity and non-balanced diets. These habits lead to metabolically and/or physiologically change that increase the risk of suffering NCD: hypertension, overweight/obesity, hyperglycemia and hyperlipidemia. The higher costs in the long term treatment of NCDs determine a region with high rates of extreme poverty such as in a developing country. WHO[3] has been developing a various types of strategies in order to reduce risks factors that include early detection and on-time treatment of diseases.

One of the characteristics of OWL-DL is its OWA[4] and how it affects to axioms defined in any ontology. As a result, those axioms must be clearly defined in order to avoid inconsistencies on reasoning time. These axioms involve several classes, then

---

[1] Non-Communicable Diseases
[2] Semantic Web Rule Language
[3] World Health Organization
[4] Open World Assumption

the reasoner will properly deduct their child classes and thus produce more accurate results.

In Section 2, the related work about spatial and biomedical ontologies is presented. The ontology construction methodology is discussed in Section3. In Section 4, the SWRL Rules are properly presented considering their application in the lifestyle patterns deduction. Finally, in Section 5 the conclusions and future work are presented.

2 Previous Work

Ontology integration can be achieved in three main ways: by merging ontologies, by mapping local ontologies to a global ontology, and by integrating local ontologies by means of semantic bridges that define mappings between the ontologies. Ontology merging is suitable for use in traditional systems which are small or moderate in size and are fairly static, and where scalability is not a core requirement. In ontology mapping, specific ontologies can be derived from global or 'reference' ontology. Ontology mapping in this case becomes much easier since concepts in different ontologies that need to be mapped are derived from the same ontology.

Our currently work on ontology integration is based on a new approach of interaction between ontologies [1, 2] which is called ontology system, where a set of ontological modules with semantic relationships among them are defined. This process allows specifying domain ontologies separately and then integrating them into a new ontology, where rules are defined to generate new knowledge. It is necessary to note that based from another work introduced by Rodriguez, J., Romero, M., & Bravo M., the methodology design is carried out successfully incorporating the elicitation term i.e., the competence questions and the division in phases of the methodology.

The last methodology allows for a seamless integration among distinct individual ontologies as is shown in [3] and [4]. This has been a first approach to achieve health & spatial ontology integration.

There exist works in the mapping health ontologies area, where the systems like ICD-10[5] or RxNorm[6] are discussed. Both ICD-10 and RxNorm are vocabularies of diseases, symptoms and findings. The first is an WHO's effort and the second one is a work from the USA´s Health organization. Another relevant approach, Ontology Integration Systems (OIS), involves descriptive logic, as a deduction mechanism. Therefore, in health ontology mapping approach exists 3 different categories: global view to local view, semantic mapping between targets and entities and mapping to enable ontology re-use [5].

In spatial ontologies field, different works have been reported using various approaches, involving the development of new spatial relationships, like spatial relationships in 3D [6]. However, it is difficult to properly define the geographic ontologies in terms of semantic, geometric and topological relationships.

A process to extend an existent geographic ontology with data mining is reviewed in [7]. Some extensions to be highlighted is concern of the semantic bridge axioms

[5] International Statistical Classification of Diseases and Related Problems 10th Revision

[6] Bodenreider, O.: Unified Medical Language system (UMLS)

layer which interacts between two ontologies, and also perform spatial reasoning by *PelletSpatial*.

Additionally, in [8], is presented another approach to join up spatial and biomedical ontologies through a bridge layer axiom which provides the necessary axioms to work between a spatial and a biomedical ontology. This work intends to reunite all the reviewed work by adding another useful approach by reusing ontologies and, implementing SWRL Rules.

3 Ontology System Development

As presented, in Section 2, the general methodology is in Bravo, M. (2014), and it is consists in the system design in 3 stages which are drawn in Fig. 1.

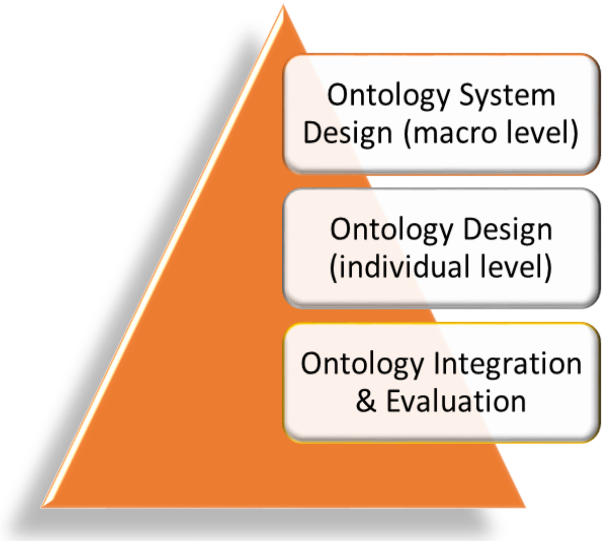


Fig. 1. Construction Methodology of a General System Ontologies.

3.1 Ontology System Design

The system integrates 6 different ontologies: *Person, NCD, Physical Activity, Nutrition, OntoMex* and *Symptom*. The integration is carried out by establishing relationships among instances of classes in different ontologies, in Fig. 2, the system's conceptual and relationship design is presented. In the case of the Region, Entity and Locality concepts, only a part of all *OntoMex* was extracted, in order to use reduced resources. To achieve the semantic relationship' conceptual design, the following competence questions were identified:

- What are the principal risk factors to suffer diabetes?
- Is the risk of suffer diabetes related with the nutrition high in fat?
- How is a mid-level lifestyle pattern defined?
- What are the characteristics of a person with nutrition high in fat?

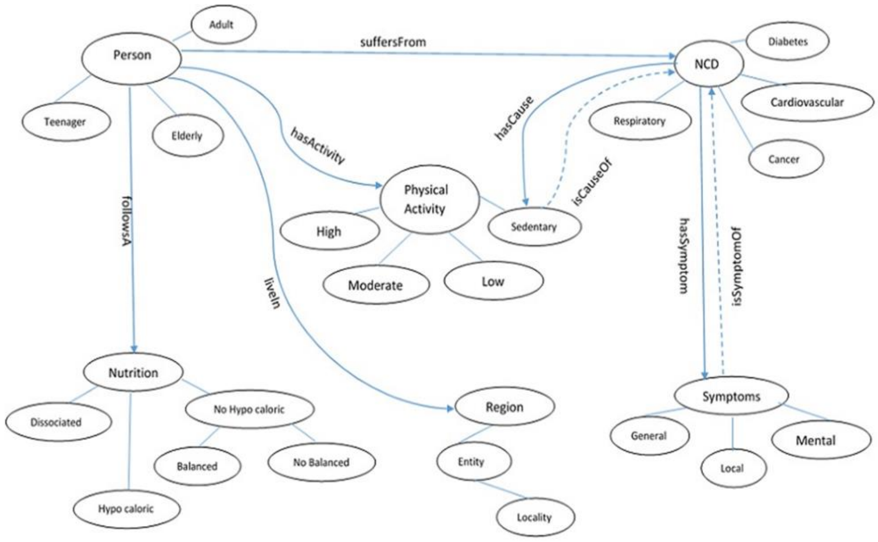


Fig. 2. Ontologies system relationships.

3.2 Individual Level Ontologies Design

Once the design stage is achieved, the ontologies which have been described in stage 1 and will form part of the system will be designed.

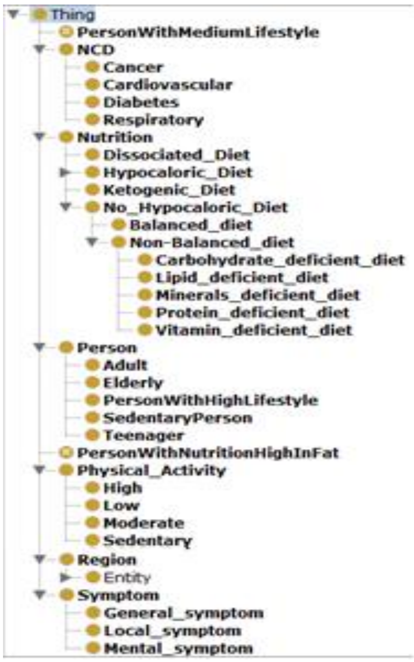


Fig. 3. Lifestyle Pattern Ontology.

In Fig. 3, these ontologies are presented as classes, once imported by the Protége's Import Ontologies tool. As an test of the system, in order to check the consistency of SWRL Rules, the OntoMex ontology have been divided, only a part of it is actually in use. The aforementioned extracting process was made to avoid overflow memory.

In Fig. 4, the ontologies system's graph is shown. Note that this diagram is not complete yet, and in subsequent works the specialization in some ontologies will be make.

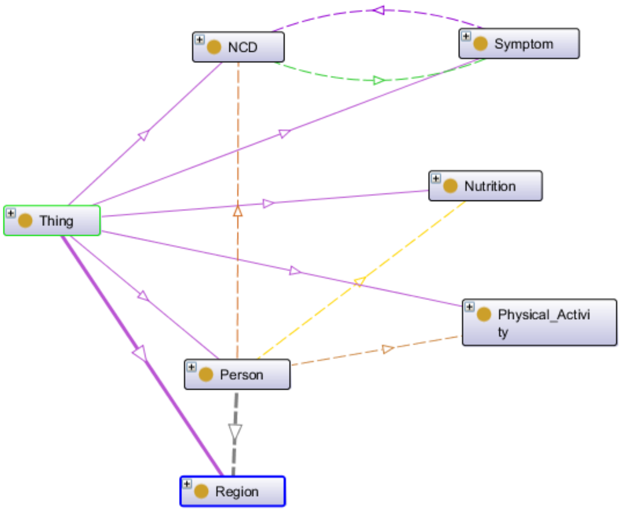


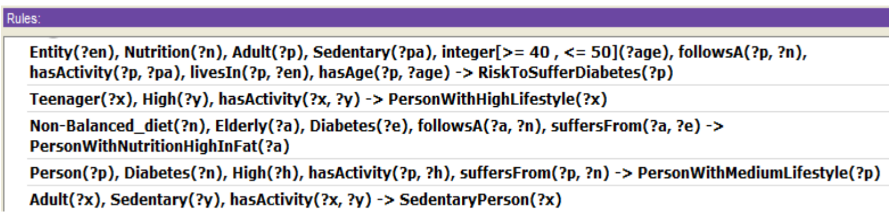
Fig. 4. Graph of the system's classes in Protégé.

3.3 Integration & Evaluation

Finally the Integration & Evaluation Stage is defined. This stage will be performed by using the SWRL Rules in order to deduct the pattern lifestyles. First of all, the axioms which are operating in the system will be designed and implemented. Those axioms are the following:

- Datatype properties,
- Object properties,
- Classes needing to be inferred,
- Necessary & sufficient conditions.

The SWRL Rules which are being used are shown in Fig. 5.



Rules:

Entity(?en), Nutrition(?n), Adult(?p), Sedentary(?pa), integer[>= 40 , <= 50](?age), followsA(?p, ?n), hasActivity(?p, ?pa), livesIn(?p, ?en), hasAge(?p, ?age) -> RiskToSufferDiabetes(?p)

Teenager(?x), High(?y), hasActivity(?x, ?y) -> PersonWithHighLifestyle(?x)

Non-Balanced_diet(?n), Elderly(?a), Diabetes(?e), followsA(?a, ?n), suffersFrom(?a, ?e) -> PersonWithNutritionHighInFat(?a)

Person(?p), Diabetes(?n), High(?h), hasActivity(?p, ?h), suffersFrom(?p, ?n) -> PersonWithMediumLifestyle(?p)

Adult(?x), Sedentary(?y), hasActivity(?x, ?y) -> SedentaryPerson(?x)

Fig. 5. Ontologies system's SWRL Rules.

In actual state of the system, SWRL rules have been checked with 85 individuals and an automatic insertion from public datasets have been in research.

As was being described, the *datatype properties axioms* or restrictions were used, then, a rule like (1), classifies a person if is an adolescent by means of their age range:

Person(?p), integer[>=12, <=17](?age), hasAge(?p, ?age) -> Teenager(?p). (1)

Datatype properties are not enough to express all the capabilities of systems relationships', i.e., accessing to the other ontologies (classes) using or design and implement another rule like (1). As a result, *Object Properties* were the next phase in the system axiomatization.

Object Properties and its proper definition of classes range and domains, guarantees the integration in a more specific level. Rule 2 was defined in order to verify the statement, and it expresses that "*If a person has a little or none physical activity, then is a sedentary person*".

Person(?p), Physical_Activity(?act), hasActivity(?p, ?act) -> SedentaryPerson(?p). (2)

SedentaryPerson have been defined and declared as a Person's subclass. Then, *SedentaryPerson* will be an Inferred Class when running the reasoner and it will contain the individuals fulfilling these restrictions.

The other Inferred Class at this stage will be *PersonWithNutritionHighInFat*, and its correspondent rule is shown in (3), which classifies an Elderly Person with high fat consumption and suffering Diabetes.

Non-Balanced_Diet(?n), Elderly(?a), Diabetes(?h),followsA(?a, ?n), suffersFrom(?a, ?e) -> PersonWithNutritionHighInFat (?a). (3)

4 Deduction by applying SWRL Rules

The mid-level lifestyle is founded by the rule in (4) which states "if a person suffers from diabetes, but have a high physical activity, then their lifestyle can be mid-level". In Fig. 6 an example of the mid-level pattern is shown.

Person(?p), Diabetes(?n), High(?h),hasActivity(?p, ?h), suffersFrom(?p, ?n) -> PersonWithMediumLifestyle(?p). (4)

The following lifestyle pattern is called the *RiskToSufferDiabetes*, as it's shown in Fig.7. It is founded by the application of rule 5 which classifies a person with a risk to suffer Diabetes. Knowledge required in rule 5 is about its physical activity which turns out to be a sedentary one, age range plus 40, living in a particular Country's State (Entity), a northern State as Nuevo Leon as an example.

Entity(?en), Nutrition(?n), Adult(?p), Sedentary(?pa), integer[>=40, <=50](age), followsA(?p, ?n), hasActivity(?p, ?pa), livesIn(?p, ?en), has Age(?p,?age) -> RiskToSufferDiabetes(?p). (5)

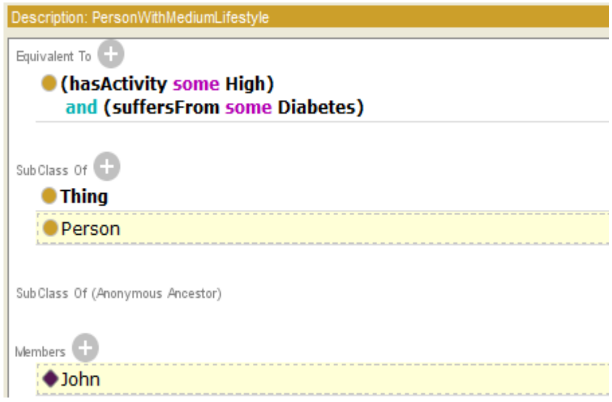


Fig. 6. Medium Lifestyle Pattern´s example.

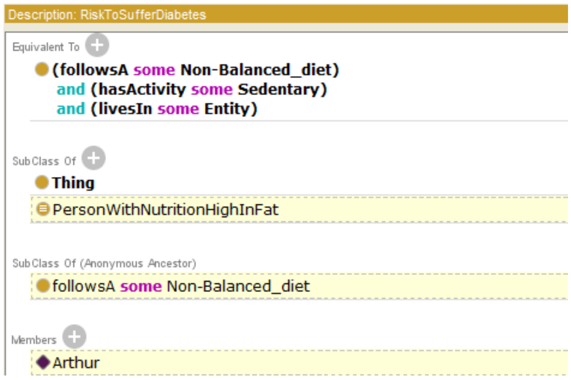


Fig. 7. RiskToSufferDiabetes Lifestyle Pattern.

As a consequence of *RiskToSufferDiabetes* pattern, *RiskToSufferDisease* can be deducted by interchanging certain parameters to concur the value partitions with, and the result is shown in Fig. 8. The pattern it's founded by rule (6), which in natural language can be expressed as: classify a person with a high level of risk to suffer a disease, if this person has a sedentary physical activity and age about plus 40.

High_Level(?hl), Person(?n), Sedentary(?pa), hasActivity(?p, ?pa),
hasLevelOfRisk(?s, ?lr) -> PersonWithHighRiskToSufferDiabetes(?p). (6)

It has High Level of values of LevelOfRisk value Partition.

Finally in Fig. 9, a later reasoner' run it is shown, denoting *RiskToSufferDiabetes* class inferred as subclass of *PersonWithNutritionHighInFat*. This statement is obvious in a normal human knowledge and reasoning. However, OWL-DL[7] 2 is based in OWA, which states that a statement is not false until it is completely proven to be false, i.e., it must have been clearly expressed whatever class A is a subclass of another class B, by means of some restriction that clearly implied the fact. If such restrictions were not correctly defined or did not exist, the reasoner will infer the

[7] Web Ontology Language based in Description Logic

information not complete; therefore, class A could be subclass of another class C, instead of class B.

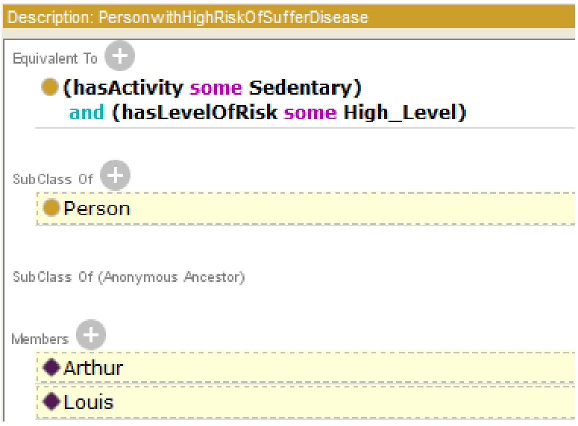


Fig. 8. RiskToSufferDisease Lifestyle Pattern.

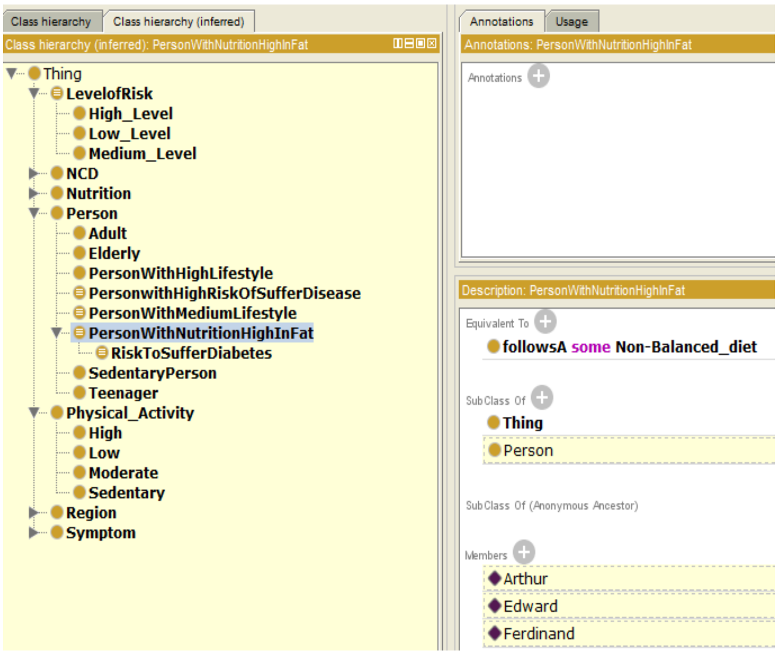


Fig. 9. Reasoner running with RiskToSufferDiabetes as a subclass.

5 Conclusions

The lifestyle patterns found through the SWRL rules application have been useful in discovering relationships among multiple dimensions of information such as: people, symptoms, NCDs, physical activity and geographical regions.

The ontology system is useful to make decisions about multiple data, all related and contained in different ontologies. Also, it shows the descriptive logic rules can be able to response the competence questions resulting from the analysis of requirements. Defining necessary and sufficient conditions' axioms allow inferred results showing more accuracy. As an example, it means to infer correctly classes as child classes or subclasses of other classes with similar restrictions. This particularly obeys to characteristics of individuals which are member of the aforementioned classes.

Ongoing work is about to migrate the current system to a server approach. Under this platform OntoMex could be loaded entirely and SWRL Rules could also be checked with all the individuals. This will allow inclusion of certain localities which could not be added for storage costs reasons.

Finally, other specializations are being planned related to the nutrition and symptom ontologies. Such that specializations permit more precise scenarios when dealing with specific diseases in order to have inferred diagnostics closer to actual cases.

Acknowledgments. We thank the Vice-Rector for *Docencia* at the Autonomous University of Puebla and the National Council for Science and Technology for the support offered to perform this work.

References

1. Bravo, M., Rodriguez, J., Pascual, J.: SDWS - Semantic Description of WebServices. International Journal of Web Services Research 11(2) (2014)
2. Bravo, M.: Similarity Measures for Web Service Composition Models. International Journal on Web Services Computing 5(1) (2014)
3. Somodevilla, M., Mena, I., Pineda, I.H., Perez de Celis, M.C.: Deducting Lifestyle Patterns by Ontologies' SWRL Rules. In: Spies, M., Wagner, R.W., Tjoa, A.M. (eds), 26th International Workshop on Database and Expert Systems: DEXA 2015, IEEE CPS (2015)
4. Somodevilla, M., Perez de Celis, M.C., Pineda, I.H., Colmenares, L.E., Mena, I.: System Development Ontology to Discover Lifestyle Patterns Associated with NCD. In: Ortuño, F. and Rojas, I. (eds), IWBBIO 2015, LNCS, Part II, Vol. 9044, pp. 47–56, Springer International Publishing Switzerland (2015)
5. Puri, C., Gomadan, K., Jain, P., Yeh, P.Z., Verma, K.: Multiple Ontologies inhealthcare InformationTechnology: Motivations and Recommendation for Ontology Mapping. In: International Conference on Biomedical Ontologies, pp. 367–369 (2011)
6. Laurini, R.: Geographic and spatial relationships. In: Input 2012 (2012)
7. Somodevilla, M., Perez de Celis, M.C., Hernandez, J., Pineda, I.H., Colmenares, L.E.: GeoHealthOntoMex: Extending a Geographic Ontology using Data Mining. In: ENC 2014, (2014)
8. Somodevilla, M., Perez de Celis, M.C., Pineda, I.H., Hernandez, J., Carrillo, M., Mena, I.: Development of an Ontologies System for Spatial Biomedical Applications. In: MICAI 2014, pp. 24–28 (2014)
9. Rodriguez, J., Romero M., Bravo, M.: Dynamic Context for Document Search and Recovery. In: Murgante, B., Misra, S., Carlini, M., Torre, C., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) Computational Science and its Applications, pp. 452–663. Springer Berlin Heidelberg (2013)

10. Bravo, M.: Ontology to Represent Similarity Relations between Public Web Services. In: Meersman, R., Dillon, T., Herrero, P., (eds.). On the Move to Meaningful Internet Systems: OTM 2011 Workshops, LNCS, Vol. 7046, pp.434–443, Springer Berlin Heidelberg (2011)
11. Donnelly, M., Bittner, T., Rosse, C.: A formal Theory for spatial representation and reasoning in biomedical ontologies. Art. Int. in Med. 36(1), pp. 1–27 (2006)
12. Smith, B., Jose, L.V., Mejino, Jr., Schulz, S., Kumar, A., Rosse, C.: Anatomical Information Science. In: Cohn, A.G. and Mark, D.M. (eds), COSIT 2005. LNCS, Vol. 3693, pp. 149--164, Springer-Verlag Berlin Heidelberg (2005)
13. Donnelly, M., Bittner, T.: Spatial Relations between Classes of Individuals. In: Cohn, A.G. and Mark, D.M. (eds), COSIT 2005. LNCS, Vol. 3693, pp. 182–199, Springer-Verlag Berlin Heidelberg (2005)
14. Fielding, J., Marwede, D.: The Image as Spatial Region: Location and Adjacency within the Radiological Image. In: Bennett, B. and Fellbaum, C. (eds), Fourth International Conference FOIS 2006, IOS Press, pp. 89–100 (2006)