

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Cervical Cytology Smear Image Interpretation System Based on Semantic Reasoning

Xu Chuanyun and Zhang Yang

College of Computer Science and Engineering, Chongqing University of Technology, Chongqing, China
College of Computer and Information Science, Chongqing Normal University, Chongqing, China

Abstract: Based on the simulation to the fact that the cytological physician's interpretation of cervical smears could be seen as the analysis and reasoning process of smear images, an Automatic Interpretation Method of Cervical Smears based on Ontology and Semantic Reasoning (AICSOSR) is put forward. Firstly the principle and process of interpretation methods using semantic reasoning is analyzed; secondly smear image features ontology, smear cytological feature ontology and interpretation experience standard ontology are built with semantic mapping methods between the different semantic ontology using rule inference and basic reasoning rules are brought forward; an automatic interpretation method of cervical smears based on Ontology and semantic reasoning is given to elaborate basic methods of interpretation rules, interpretation result merging and conflict management.

Key words: Cervical smear, ontology, semantic logic reasoning

INTRODUCTION

Automatic cervical smear interpretation method based on ontology and semantic logical reasoning (SORSI) is to use ontology and description logics to do standard formal definition with some concepts such as image feature, cytological characteristics, cervical lesions and then to extract features from smear images by using image processing techniques and then to map image features for cytological characteristics by using semantic logic reasoning and finally to get interpretation results by use of logical reasoning according to the relationship between cervical lesion and cytological characteristics. Semantic reasoning interpretation method is inspired by the artificial interpretation. The process of cytological physician to observe cervical smears: (1) By scanning smears under the microscope, the visual system receives cell images from the microscopic; (2) The recognition of image would be got based on the normal knowledge and medical knowledge as well as done with concept description and then the cytological features of cervical lesions (e.g., kernel partial large cells, eosinophils) which may be from cytology physicians' logic understanding to the microscope image would be searched in the brain and those understanding should be based on cytology physicians' common sense, medical knowledge, microscopic image visual features and (3) Combining with the interpretation knowledge (interpretation standard), the interpretation results would be obtained by logically reasoning.

The interpretation model in this study is made up of three main parts:

- **Image processing system:** To construct the image features by digitizing smear
- **Medical knowledge base (MKB):** Including the characteristics description ontology and reasoning rules
- **Logic reasoning system:** To formally represent image features, perform inference rules, generate the cytological features of smears, synthesize cytological characteristics and implement interpretation rules

ONTOLOGY MODEL FOR CERVICAL SMEAR

Cervical smear interpretation domain ontology is the foundation of interpretation inference of automatic interpretation system as well as the glossary with a clear semantics which can be jointly understood by both human and computer and involved in the interpretation process. In this study, the ontology is divided into four categories which are from the underlying semantic to high-level semantic in turn-image feature ontology, cytological features ontology, interpretation results ontology, interpretation experience standard ontology and the relationship between the ontology is shown in Fig. 1.

The semantics of the top ontology is constructed by the lower ontology, for example the dyeing depth of smear elements ontology is made up from rule inferences of the

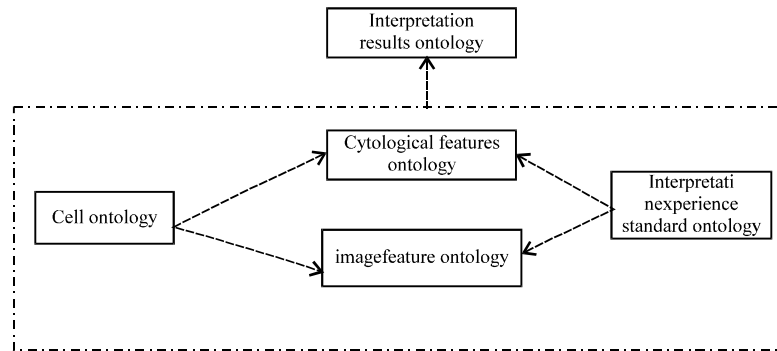


Fig. 1: Cervical smears hierarchical semantic model

color depth of the nucleus region in the smear image area ontology. The build process from the bottom up is a map of ontology semantics and this study realizes this mapping from low-level to high-level semantic meaning by rules reasoning.

Ontology modeling of cytological features: Cytological features ontology is semantic concepts involved in understanding smears in interpretation and the semantic basis and semantic representation of interpretation. Cytological characteristics ontology is divided into two main categories: the one is the cell type ontology such as normal cells, abnormal cells; the other one is the ontology on behalf of the cytological features such as the size of cells, chromatin depth. Smear elements ontology is closer to the people in the semantic cognition whose concrete semantics would be mapped from image features ontology.

Cell type ontology is the classification of cell interpretation and mainly divided into two kinds of normal cells and abnormal cells. Normal cells (cfo:Normal) includes the columnar cells (cfo: Columnar), intermediate cells (cfo: Intermediate), surface cells (cfo: Superficial) as well as abnormal cells (cfo:Abnormal) includes mild squamous intraepithelial lesion cells (cfo: LightDysplastic), moderate squamous intraepithelial lesion cells (cfo: ModerateDysplastic), high grade squamous intraepithelial lesion cells (cfo: SevereDysplastic), squamous cell carcinoma cells (cfo: CarcinomaInSitu) and all the cells are the sub class of cell body (cfo:Cell). The relationship of various cell type ontologies is shown in Fig. 2. In the cell type ontology, normal cells are known as negative cells as well as abnormal cells are known as positive cells.

The cytological feature ontology is the semantic description of cells features and the semantic basis of cervical smears interpretation. Ontology modeling of smear cytology features is based on the knowledge of

cytology and physicians' perceptions to smears. All cytological features are inherited from the same upper ontology as well as each cytological feature ontology indicates one kind of cytological characteristics and the concrete value of cytological characteristics is expressed by individuals corresponding to the ontology. The inheritance relationship between smear cytological feature ontology is shown in Fig. 3.

Ontology modeling of interpretation experience standards: Browsing smears through the microscope, the cytological physician could form the preliminary judgment of cells in smears in a very short time. If the preliminary determination may have lesions, the cytological physician could stay in sight for the careful observation of cells and sometimes even need to adjust the magnification of microscope to observe carefully at high magnification. In order to simulate human's interpretation based on experience, this study counts characteristics of images by use of different types of cells learning samples, sets range for different image features in different cells and the image features in the set range could be thought to be associated with this species of cells.

Each cell in smears have corresponding interpretation experience standard as well as all experience interpretation standard inherit from cfo: CellTypeStd and Fig. 4 shows the inheritance relationship of interpretation experience standard ontology.

The results of interpretation experience standard results have three kinds: assurance (cfo: FeatureInterpretState_confidence), doubt (cfo: FeatureInterpretState_suspect) and irrelevance (cfo: FeatureInterpretState_unrelated) which denote the coincidence degree of the interpreted cells with and interpretation experience standard in the image features. Four associated range values corresponding with the above interpretation results are the assurance upper limit value (cfo: hasConfidenceIntervaHigh) and the assurance

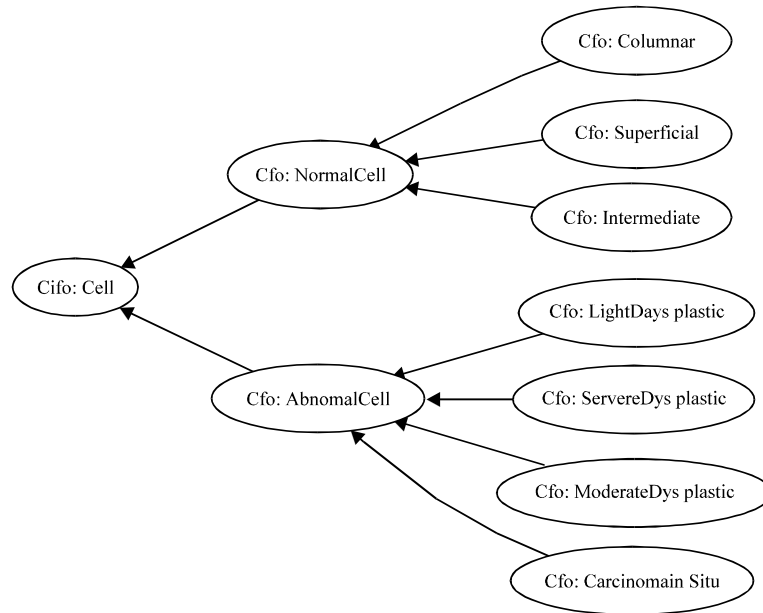


Fig. 2: Relationship between cell types of ontology

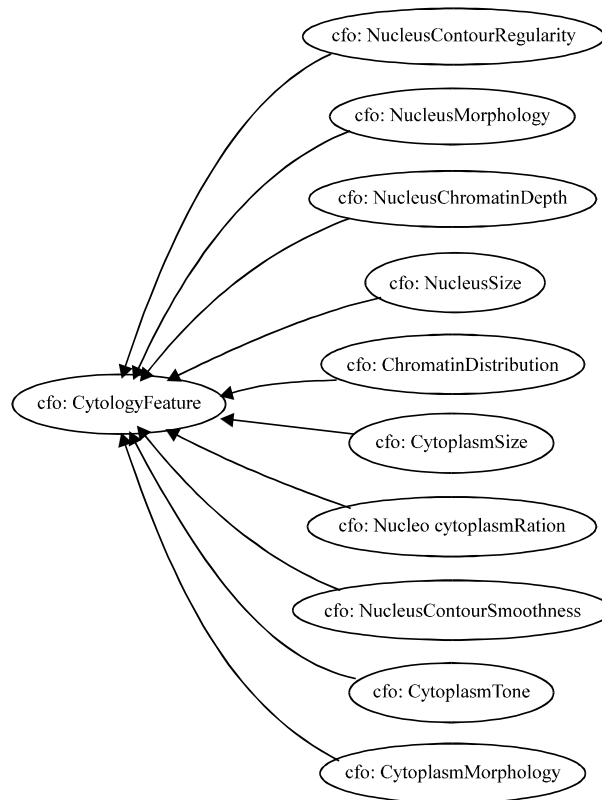


Fig. 3: Ontologies of cytological features

lower limit value (cfo: hasConfidenceIntervalLow), the doubt upper limit value (cfo: hasSuspectIntervalHigh) and

the doubt lower limit value (cfo: hasSuspectIntervalLow), these four values are text attributes of interpretation

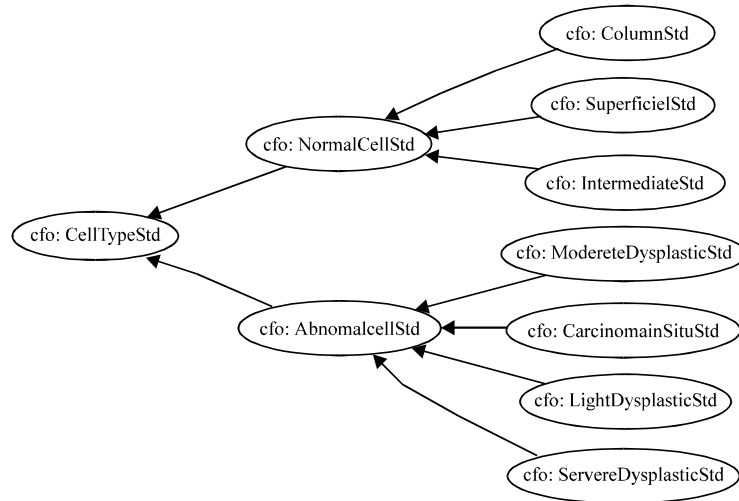


Fig. 4: Associated with the interpretation of the empirical standard ontology inheritance

experiences calibration ontology (cfo:InterpreStd). In addition to these four attributes, are text attributes of ontology (cfo:InterpreStd) also include the mean (cfo: HasMeanValue) and standard deviation (cfo: HasStdValue).

Each interpretation standard sets the corresponding experience scope for each image feature, for example the interpretation experience standard ontology (cfo:AbnormalStd) of abnormal cells has object properties (cfo: hasNucleocytoplasmicRatioStd) with value range (cfo:InterpreStd) which denote the interpretation experience standard of nuclear-cytoplasmic ratio.

SEMANTIC MAPPING OF CERVICAL SMEAR ONTOLOGY

In our AICSOSR, there should be two cell image features and cytological characteristics which are associated with each other as well as not in the same semantic level. The image features are low-level features to describe image properties of smears which mainly use the terminology related to image processing; cytological characteristics are semantic interpretation oriented and direct evidence to interpretation which belong to high-level semantic features and mainly use cytological terms. Cytological features of smears come from image features of smears and image features directly extracted from images by image analysis techniques. In this study, these two features would be formally defined by ontology to get image feature ontology and smear cytological features. Because two kinds of feature ontology are defined to smear attributes from different angles, the semantic mapping of characteristics could be converted into Ontology mapping problem.

Ontology mapping is a process to take two ontologies in which there is some concept association of semantic level as input and establish the corresponding semantic relations for the various elements (concepts, relationship, instances) of two ontologies (Yang, 2011). Ontology mapping technology is divided into: (1) The element level ontology mapping technology including the technology based on strings (Cohen *et al.*, 2003), technology based on language (Marynard and Ananiadou, 1999), technology based on constraints (Doan and Halevy, 2005), technology based on linguistics (Miller, 1995), mapping results reuse technology (Do and Rahm, 2002; Aumüller *et al.*, 2005); (2) the structure level ontology mapping technology including technology based on diagram (Rahm *et al.*, 2004), technology based on classification system (Garey and Johnson, 1979), technology based on structure library (Petko and Euzenat, 1997), technology based on model (Bouquet *et al.*, 2006), data analysis and statistical techniques (Anhai *et al.*, 2002; Prasad *et al.*, 2002). This study uses semantic reasoning method based on rules to map image characteristics ontology to smear cytological characteristics ontology.

The automatic interpretation method of cervical smears based on Ontology and semantic reasoning takes as the knowledge base and semantic reasoning core and its system model is shown in Fig. 5. The model is formed by five main modules – Smear scanning, Image processing, Feature mapping, Individual cells interpretation, Interpretation merging and conflict management as well as three auxiliary modules- Interpretation signs, Artificial feature annotation and interpretation, Annotation results comparing:

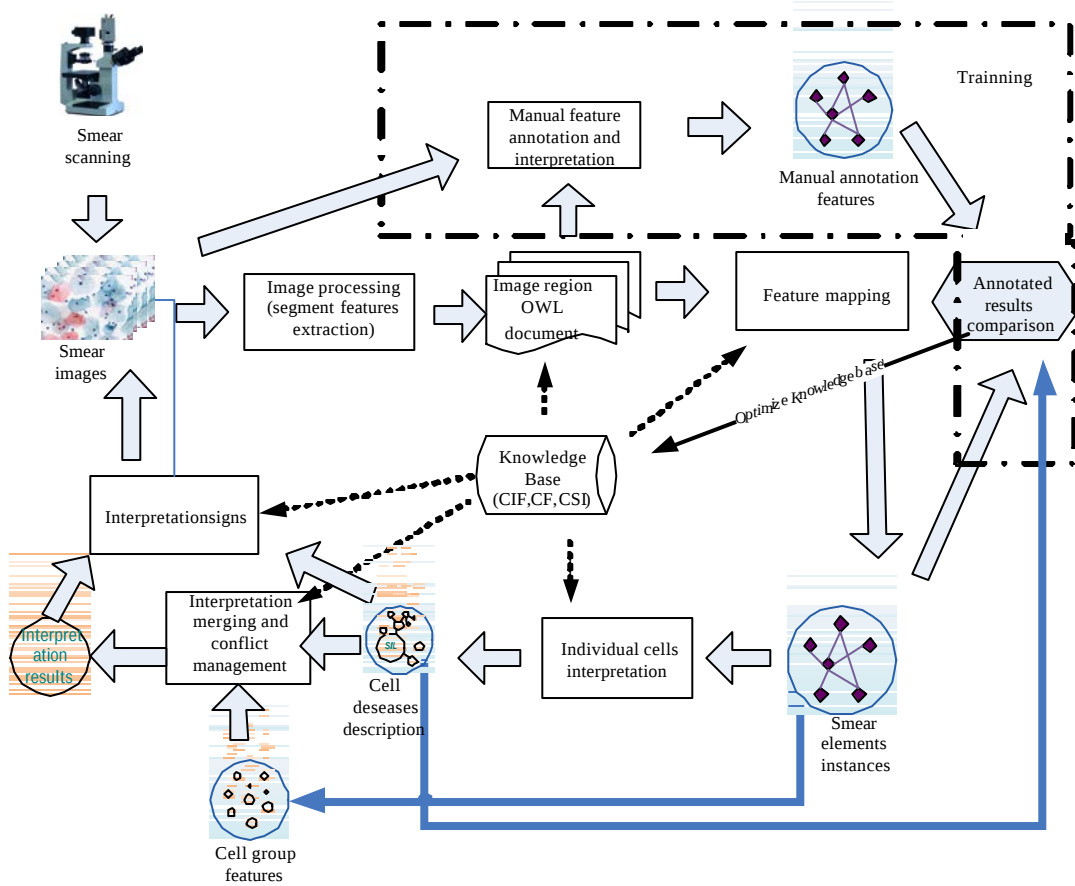


Fig. 5: Interpretation system model

- **Smear scanning:** Digitizes smears by using the slide scanning system to get digital slide (digital images)
- **Image processing:** Segments cytological picture elements from the background to form independent image area and extracts morphological features, texture features, color features for each area and finally maps every image area to ontology stored by OWL format. This study uses the cell image feature to describe image features of area
- **Feature mapping:** maps image features to cytological features to form advanced cytological semantic representation
- **Individual cells interpretation:** analyzes cytological features of each cell and makes lesions interpretation on each cell by using semantic rules inferences according to cytological characteristics for each cervical lesion
- **Interpretation merging and conflict management:** comprehensively analyzes some elements such as interpretation of single cells, the characteristics of cell group and individual patient characteristics to

make the final interpretation. The conflict management is to eliminate conflicts and make clear reliable interpretation by integrating a variety of factors when a variety of features point to different interpretation results

- **Interpretation signs:** visually displays interpretation process and interpretation results in smears images and identifies image feature, cytological features and interpretation results in each area

The interpretation method of individual cells is to formulate corresponding interpretation rules for each interpretation result and judge the one to be the corresponding interpretation results if it meets the rules. For example, the interpretation rules of low grade squamous intraepithelial lesion (LSIL) are as follows:

- `#interpret_cell_lsil`
- `cfo:Cell (?c) ∧ cfo:hasCytoplasm (?c, ?cc) ∧ cfo:hasNucleus (?c, ?n) ∧`
- `cfo:hasCytoplasmSize (?cc, ?cs) °`

- sqwrl:makeBag (?cb, cfo:cytoplasmSize_larger) ^ sqwrl:element (?cs, ?cb) ^
- cfo:hasNucleusSize (?n, ?ns) ^
- sqwrl:makeBag (?nb, cfo:nucleusSize_larger) ^
- sqwrl:element (?ns, ?nb) ^
- cfo:hasNucleocytoplasmicRatio (?c, ?ncr) ^
- sameAs (?ncr, cfo:nucleocytoplasmicRatio_high) ^
- cfo:hasNucleusChromatinDepth (?n, ?ncd) ^
- sqwrl:makeBag (?ncb, cfo:nucleusChromatinDepth_LightlyHyperchromatic) ^
- sqwrl:element (?ncd, ?ncb) ^
- cfo:hasChromatinDistribution (?n, ?nd) ^
- sqwrl:makeBag (?ndb, cfo:chromatinDistribution_normal) ^
- sqwrl:element (?nd, ?ndb) ^
- cfo:hasNucleusContourRegularity (?n, ?ncr1) ^
- sqwrl:makeBag (?ncrb1, cfo:nucleusContourRegularity_normal) ^
- csi:isLSIL (?c, csi:lsil)

SWRL inference rules are based on the open world assumption but the combination of above rules are based on the closed world assumption, which means that it cannot be interpreted to be LSIL, HSIL, KSCC and NKSCC according to the rules above. The interpretation based on the close world assumption is based on TBS interpretation standards whose basic principle is that we can think it no lesion if the characteristics of lesions are not found. Therefore interpretation methods based on closed world assumption would have risks about pathological changes or missing because there is no criterion rule for no intraepithelial lesion and malignant lesion (NILM).

EXPERIMENT AND ANALYSIS

In order to validate the feasibility and effectiveness of automatic interpretation method of cervical smears based on Ontology and semantic reasoning, this study chose cervical cell images in Herlev (Jantzen *et al.*, 2005) cervical cell image data set as experimental data. The experiment used Matlab to make image preprocessing at first and used the method in this study to make feature extraction and then randomly selected one half of the samples to make interpretation rules learning as well as the other half to do experiments as the interpretation data.

The experiment used the nucleo cytoplasmic ratio and nuclear size as the main reference for interpretation and got the results shown in the Table 1.

In the Table 1, the class normal_columnar in the positive interpretation results was not positive cells which belonged to false positive interpretation; the class

Table 1: Interpretation results

Interpretation results	Cell types	
Positive interpretation	carcinoma_in_situ	75
	light_dysplastic	90
	moderate_dysplastic	73
	normal_columnar	18
	severe_dysplastic	96
Negative interpretation	Totality	352
	light_dysplastic	1
	normal_columnar	7
	normal_intermediate	35
	normal_superficial	61
	severe_dysplastic	3
	Totality	107

Table 2: Analysis about interpretation results

Cell No.	True	False	Totality	FN(%)	FP(%)	OE(%)
Negative	103	4	121	1.18		
Positive	334	18	338		14.88	
Totality	437	22	459			4.80

light_dysplastic, severe_dysplastic in the negative interpretation results was not negative cells which belonged to false negative interpretation. In the positive interpretation results, the class normal_columnar was the cells with the most error reading, because interpretation rules used in this experiment was mainly based on the nucleo cytoplasmic ratio and nuclear area as well as the nucleo cytoplasmic ratio and cell nuclear area of columnar cells were close to diseased cells.

Table 2 was the analysis of interpretation results.

Known from the Table 2, in the Herlev data set, the interpretation accurate rate on negative, positive two classifications the interpretation method in this study was 95.2% as well as the correct rate of using neural network classification in the literature (Norup, 2005) was 94.33%, which meant that the method in this study was better than neural network.

SUMMARY

Based on OWL ontology description language and SWRL rule language, this study put forward automatic interpretation method of cervical smears and establishes the corresponding system model. Firstly we build reasoning interpretation model of cervical smears and introduces the principle and every component of the model; secondly we build the hierarchical semantic model of cervical smears by analyzing the characteristics of cervical smears and labor ontology modeling process and results of each level; thirdly we put forward ontology semantic mapping method based on semantic rule inference and discuss mapping method and related inference rules image feature ontology of cell, cytoplasm, nucleus to cytological feature Ontology in detail; fourthly we construct automatic interpretation system model of cervical smear based on Ontology and semantic reasoning

and detail the principle of model and interpretation methods, interpretation rules, interpretation results merging of individual cells; finally we verify the feasibility and effectiveness of the proposed method and model by doing single cell judgment experiment on Herlev cervical cell image data set.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China under Grant No. 61202348, the National Natural Science Foundation of China under Grant No. 61173184, the Key Science and Technology Project of Chongqing under Grant No. CSTC2012GGYYJS10027 and the Natural Science Foundation Project of Chongqing under Grant No. CSTC2012jjA1549.

REFERENCES

- Anhai, D., J. Madhavan and P. Domingos, 2002. Learning to map between ontologies on the semantic web. *Proceedings of the 11th International Conference on World Wide Web*, May 7-11, 2002, Honolulu, Hawaii, USA., pp: 662-673.
- Bouquet, P., L. Serafini and S. Zanobinik, 2006. Bootstrapping semantics on the web: Meaning elicitation from schemas. *Proceedings of the 15th International Conference on World Wide Web*, May 23-26, 2006, Edinburgh, Scotland, pp: 505-512.
- Cohen, W.W., P. Ravikumar and S.E. Fienberg, 2003. A comparison of string metrics for matching names and records. *Proceedings of the KDD Workshop on Data Cleaning and Object Consolidation*, August 2003, Washington, DC., pp: 73-77.
- Aumuellner, D., H.H. Do, S. Massmann and E. Rahm, 2005. Schema and ontology matching with COMA++. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, June 14-16, 2005, Baltimore, Maryland, USA., pp: 906-908.
- Doan, A. and A. Halevy, 2005. Semantic integration research in the database community. *AI Mag. Special Issue Semantic Integration*, 26: 83-94.
- Do, H.H. and E. Rahm, 2002. COMA-a system for flexible combination of schema matching approach. *Proceedings of the 28th VLDB Conference*, August 20-23, 2002, Hong Kong, China, pp: 610-620.
- Garey, M.R. and D.S. Johnson, 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, USA.
- Jantzen, J., J. Norup, G. Dounias and B. Bjerregaard, 2005. Pap-smear benchmark data for pattern classification. *Proceedings of the Nature Inspired Smart Information Systems*, October 3-5, 2005, Albufeira, Portugal, pp: 1-9.
- Marynard, D. and S. Ananiadou, 1999. Term Extraction Using a Similarity-Based Approach. In: *Recent Advances in Computational Terminology*, Meyer, I., K. Mackintosh, C. Barriere and T. Morgan (Eds.). John Benjamins Publishing Co., Amsterdam, Philadelphia, pp: 261-278.
- Miller, G.A., 1995. WordNet: A lexical database for English. *Commun. ACM*, 38: 39-41.
- Norup, J., 2005. Classification of pap-smear data by transductive neuro-fuzzy methods. Masters Thesis, Technical University of Denmark, Oersted-DTU.
- Petko, V. and J. Euzenat, 1997. Dissimilarity measure for collections of objects and values. *Proceedings of the 2nd Symposium on Intelligent Data Analysis*, August 4-6, 1997, London, UK., pp: 259-272.
- Prasad, S., Y. Peng and T. Finin, 2002. A tool for mapping between two ontologies using explicit information. *Proceedings of the AAMAS Workshop on Ontologies and Agent Systems*, July 07, 2002, Bologna, Italy.
- Rahm, E., H.H. Do and S. MaBmann, 2004. Matching large XML schemas. *ACM SIGMOD Record*, 33: 26-31.
- Yang, F., 2011. Research on key technologies of ontology mapping. Jilin University, China.