

## A Study of Natural Language Processing Procedures

M. Muralidharan<sup>1\*</sup>, V. Valli Mayil<sup>2</sup>

<sup>1\*</sup>Department of Computer Science, Periyar Maniammai University, Vallam Thanjavur, India

<sup>2</sup>Department of Computer Science and Applications, Periyar Maniammai University, Vallam Thanjavur, India

\*Corresponding Author: [mi.murali123@yahoo.com](mailto:mi.murali123@yahoo.com)

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Received: 25/May/2017, Revised: 02/Jun/2017, Accepted: 20/Jun/2017, Published: 30/Jun/2017

**Abstract**— Informatics techniques, for example, content mining and natural language processing, are constantly required in bioinformatics examine. In this review, we talk about content mining and natural language processing techniques in bioinformatics from two points of view. Initially, we plan to scan for information on science, recover references utilizing content mining techniques, and reproduce databases. For instance, protein connections and quality ailment relationship can be mined from PubMed. At that point, we examine the uses of content mining and natural language processing systems in bioinformatics, including foreseeing protein structure and capacity, recognizing noncoding RNA. At last, various strategies and applications, and also their commitments to bioinformatics, are talked about for later use by content mining and natural language processing analysts.

**Keywords**—NLP, Bioinformatics, Textmining, Non Coding, RNA Identification.

### I. INTRODUCTION

Content mining and natural language processing allude to understanding and breaking down natural language by utilizing PC calculations and projects. It is an imperative research bearing in the application field of manmade brainpower. Look into on natural language processing and content mining has been accounted for as ahead of schedule as the rise of PCs. With consistent and broad research on machine learning and information mining calculations, existing content mining advancements have accomplished great outcomes in programmed deliberation, programmed address replying, web social system investigation, and anaphora determination.

Bioinformatics is an entomb train that rose with the advance and achievement of the Human Genome Extend. It predicts and takes care of live science issues identified with hereditary qualities by utilizing PC and factual informatics. Information stockpiling, recovery, and investigation are the key procedures in bioinformatics. The National Center for Biotechnology Information set up different databases for organic information, including succession databases for putting away DNA and protein information (e.g., dbEST and dbSNP), Online Mendelian Inheritance in Man database for putting away infection information, Gene Expression Omnibus database for putting away quality chip information, and PubMed database for putting away natural and restorative writing.

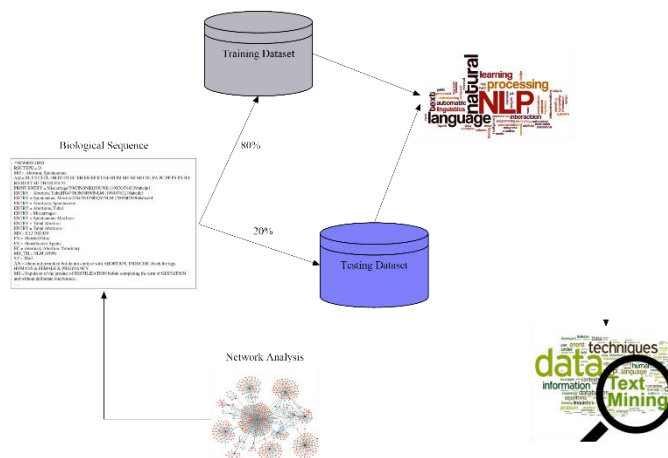


Fig.1. Complications procedure and association among NLP and bioinformatics

Content mining and natural language processing methods are important to recover client inclination learning from growing databases. Along these lines, analysts recover papers on specific points of intrigue, for example, deciding protein communications, from PubMed utilizing PC calculations and projects. With the figuring out of hereditary codes, scientists have confirmed that organic successions, especially protein arrangements, are like human language as far as structure. Notwithstanding utilizing content mining to recover bioinformatics articles straightforwardly, an expanding number of scientists are seeing protein successions as a unique "content" and breaking down them in light of existing content mining innovations.

## I. MINING BIOINFORMATICS LITERATURE

The advancement of content mining innovation assumes a vital part in recovering organic writing, especially in building up natural data databases. An extraordinary workshop on natural writing recovery issues was led amid the Annual Meeting of the Association for Computational Linguistics and the Annual International Conference on Intelligent Systems for Molecular Biology in 2005 to examine writing mining issues identified with bioinformatics. Removing protein-protein connections and the relationship between quality capacities and illnesses are two driving application subjects.

### A. *Extricating Protein-Protein Interactions.*

Separating the protein connection system is an essential research subject in bioinformatics and frameworks science. In past reviews, analysts scanned for protein-protein connections physically. Notwithstanding, with the exponential development of natural writing, a program that can perceive protein-protein collaborations consequently from PubMed digests is essential. By and by, no brought together naming guideline for proteins has been built up yet. Numerous proteins and qualities utilize a similar name. Thus, perceiving protein names from the writing abstracts and further deciding their collaborations are enter issues in the utilization of content digging in hunting down protein-protein associations.

At first, scientists separated protein-protein associations through factual and tallying strategies. They physically made word references of protein names and afterward sought modified works that include components happening at any rate twice. On this premise, analysts confirmed that related proteins interface with each other. A few analysts likewise utilized element wanting to concentrate and think about protein-protein cooperations.

Separating protein-protein collaborations has been an examination problem area in bioinformatics for quite a while and has pulled in an expanding number of scientists in the fields of content mining and natural language processing. To start with, the language structure of writing edited compositions is investigated all the more precisely, as opposed to making a straightforward insights of lexicon words. Kim et al. changed over a convoluted semantic structure examination into computing the most limited way in a diagram by making a nucleus. Comparable examination techniques for writing abstracts incorporate syntactic investigation, setting free sentence structure examination, cosmology examination, and other data recovery strategies. Protein-protein associations are analyzed utilizing these investigation techniques. What's more, many machine

realizing techniques, for example, outfit learning and Bayesian system, are connected to perceive protein names and cooperations.

### B. *Removing the Relationship between Gene Functions and Diseases.*

Removing protein-protein cooperations includes scanning for two proteins in the content and figuring out if they communicate with each other. Essentially, extricating the relationship between quality capacities and illnesses likewise includes looking for quality names and infection names all the while in the writing and afterward figuring out if a specific quality is identified with a specific malady.

When all is said in done, such extraction process can be separated into three stages. To begin with, the modified works of related papers are looked through correlation with a word reference. Second, the inquiry scope must be extended forward and in reverse once in a while in view of the area of the related word or statement to guarantee exactness. At long last, truths are assessed utilizing linguistic use examination strategies or machine learning techniques. Such extraction strategies every now and again yield great outcomes for uncommon qualities and infections. Bui et al. inspected the relationship amongst medications and HIV variety in PubMed. Jiang et al. decided the relationship between roughly 3000 microRNAs and diverse maladies in light of the naming principle of microRNA. Cheng et al. built up a content mining framework in view of the relationship among human illnesses, varieties, and medication impacts. Iossifov et al. concentrated on researching contortions of human and mouse encephalon. Jensen et al. made a point by point rundown of related record databases, writing mining programming, and capacities.

### C. *Recovering References.*

A lot of bioscience writing has been distributed. Hunting down connecting proteins and analyzing the relationship amongst qualities and maladies are just two application cases. Content mining innovation is required to get answers to numerous other bioscience and bioinformatics issues in different databases, for example, PubMed.

Organic writing mining and related critical thinking need to adapt to two noteworthy issues, in particular, perceiving name substances and removing relations. These issues are predominantly understood by (1) methods in view of phonetic analysis, (2) strategies in light of word references, (3) machine learning techniques, and (4) factual strategies.

A few critical databases are additionally chosen with content mining. STRING and BioGRID are worked for protein collaboration with writing mining. For foreseeing quality capacity, PubTator and GeneCards are essential databases utilizing content mining systems. Related works were checked on in detail in Huang and Lu's work as of late. As the improvement of group source, manufactured content seeking and mining can likewise be useful for biomedicine writing accumulation.

In addition, changing over PubMed database into an Extensible Markup Language social database [43] and a fluffy pursuit of papers and creator names through here and now coordinating are likewise flow look into problem areas.

## II. APPLYING TEXT MINING TECHNOLOGIES TO PROTEIN RESEARCH

DNA and protein arrangements are a significant hereditary language and are viewed as the fixed book of life. In this manner, an expanding number of natural language processing and content mining calculations are being connected to ponder bioinformatics. For instance, inactive semantic investigation was connected to protein remote homology location, and protein unearthly examination begins from word recurrence insights in natural language processing. Moreover, some language structure standards of protein, DNA, and RNA groupings were found, and a few web servers were built to remove these elements and tenets.

### A. Foreseeing Protein Structure.

Protein structure decides work. Henceforth, it ought to be investigated to decide protein work. The auxiliary examination of protein for the most part spotlights on certain protein successions and characterizes districts into the  $\alpha$ -helix,  $\beta$ -lamella, and protein cluttered locales. Anticipating the  $\alpha$ -helix and  $\beta$ -lamella locales is the same as foreseeing the auxiliary protein structure.

On the off chance that a protein succession is viewed as a natural language, then dissecting the kind of protein in an area is like aligning sentence structure in natural language processing. To begin with, the auxiliary protein structure is anticipated by consolidating tenets and measurements. In any case, confronted with the bottleneck of measurable expectation, a few specialists have proposed utilizing machine learning forecast techniques, including strategies in light of simulated neural system (ANN), bolster vector machine (SVM), irregular timberland, and most extreme entropy.

Anticipating the protein confused locale is likewise led. This district alludes to the region without a steady or one of a kind 3D structure in the protein space structure. Numerous content mining and machine learning techniques, including ANN, SVM, contingent arbitrary field, and irregular backwoods, have been utilized to anticipate the protein cluttered district. Regular existing server locations are recorded in Table 1.

Table 1. Web server for protein complaint estimate

Problem	Name	Websites	Input format
	DisProt	<a href="http://www.disprot.org/pondr-fit.php">http://www.disprot.org/pondr-fit.php</a>	Fasta or EMBL sequence format
		<a href="http://www.disprot.org/metapredictor.php">http://www.disprot.org/metapredictor.php</a>	
		<a href="http://www.dabi.temple.edu/disprot/predictor.php">http://www.dabi.temple.edu/disprot/predictor.php</a>	
Protein disorder prediction	DisEMBL	<a href="http://dis.embl.de/">http://dis.embl.de/</a>	SwissProt ID
		<a href="http://www.sbc.su.se/~maccallr/disorder/cgi-bin/submit.cgi">http://www.sbc.su.se/~maccallr/disorder/cgi-bin/submit.cgi</a>	Only plain sequence; one sequence once; slow
		<a href="http://bip.weizmann.ac.il/rdbin/finde">http://bip.weizmann.ac.il/rdbin/finde</a>	Only plain sequence; one sequence once
	IUPred	<a href="http://iupred.enzim.hu/">http://iupred.enzim.hu/</a>	SwissProt ID or plain sequence
		<a href="http://www.pondr.com/cgi-bin/PONDR/pondr.cgi">http://www.pondr.com/cgi-bin/PONDR/pondr.cgi</a>	Fasta
		<a href="http://bioinf.cs.ucl.ac.uk/psipred/?dispred=1">http://bioinf.cs.ucl.ac.uk/psipred/?dispred=1</a>	Raw sequence or fasta format
	SCRATCH	<a href="http://scratch.proteomics.ics.uci.edu/">http://scratch.proteomics.ics.uci.edu/</a>	Only plain sequence; one sequence once; slow
		<a href="http://distill.ucd.ie/spritz/">http://distill.ucd.ie/spritz/</a>	Raw sequence or fasta format
		<a href="http://www.strubi.ox.ac.uk/RONN/">http://www.strubi.ox.ac.uk/RONN/</a>	Fasta, but only one sequence once

### B. Foreseeing Protein Function.

Anticipating protein capacity is a standout amongst the most essential research subjects in bioinformatics. It includes anticipating protein-protein cooperations and communication destinations, limiting subcellular protein, foreseeing and grouping transmembrane protein, protein remote homology location, characterizing protein capacities, perceiving multifunctional chemicals, and DNA restricting protein ID.

The protein arrangement is anything but difficult to decide. Like natural language, the protein arrangement has many confounded principles. Notwithstanding, outlining and understanding the standards of protein successions are troublesome. In this manner, breaking down and foreseeing the "protein language" communicated by amino corrosive groupings by utilizing computational phonetics and machine learning techniques are essential. Through these methodology, we might have the capacity to comprehend the elements of protein successions.

Anticipating protein-protein communications is a standout amongst the most essential research subjects in protein capacities. Numerous specialists are focused on anticipating whether two protein groupings display collaborations. To date, many machine learning strategies have been connected, including SVM, part strategy, basic leadership tree, arbitrary woods, Bayesian system, and the autoregressive model. A few content processing techniques, for example, philosophy explanation and test weighting, are utilized to recognize elements and process preparing information. While anticipating protein-protein cooperations, specialists likewise expect to break down the locale of protein-protein connections, which is utilized to foresee protein-protein association destinations. Data approaches ordinarily utilized

as a part of linguistic examinations, for example, condition arbitrary fields and a concealed Markov demonstrate (HMM), have been utilized to dissect communication locales and have accomplished great outcomes. In addition, irregular backwoods, SVM, ANN, Bayesian system, straight relapse, and other machine learning techniques are utilized to foresee protein-protein communication destinations. All things considered, a few analysts question that deciding the protein succession alone is lacking to give adequate data to foreseeing connections. Content mining and machine learning analysts ought to grow new components and arrangement techniques to take care of this issue. The sites of existing normal programming used to anticipate protein-protein connections and association locales are given in Table 2.

Table 2. Web server for protein-protein relations and locations estimate.

Problem	Name	Websites	Input format
Protein interaction sites prediction	PPISP	<a href="http://pipe.scs.fsu.edu/ppisp.html">http://pipe.scs.fsu.edu/ppisp.html</a> <a href="http://pipe.scs.fsu.edu/meta-ppisp.html">http://pipe.scs.fsu.edu/meta-ppisp.html</a>	PDB file
	Protomot	<a href="http://protomot.csbb.ntu.edu.tw/index.html">http://protomot.csbb.ntu.edu.tw/index.html</a>	PDB ID
	SPPIDER	<a href="http://sppider.cchmc.org">http://sppider.cchmc.org</a>	PDB file or PDB ID
	Whisky	<a href="http://nmr.chem.uu.nl/Software/whisky/index.html">http://nmr.chem.uu.nl/Software/whisky/index.html</a>	PDB file
Protein-protein interaction prediction	InterPreTS	<a href="http://www.russell.embl.de/cgi-bin/tools/interprets.pl">http://www.russell.embl.de/cgi-bin/tools/interprets.pl</a>	Fasta, 40 sequences at most
	PIE	<a href="http://www.ncbi.nlm.nih.gov/CBBresearch/Wilbur/IRET/PIE/">http://www.ncbi.nlm.nih.gov/CBBresearch/Wilbur/IRET/PIE/</a>	Gene ID or name
	PPI	<a href="http://121.192.180.204:8080/PPI/Home.jsp">http://121.192.180.204:8080/PPI/Home.jsp</a>	Fasta
	PredHS	<a href="http://www.predhs.org/">http://www.predhs.org/</a>	PDB files, 10 files at most
	Pred-PPI	<a href="http://cic.scu.edu.cn/bioinformatics/predict_ppi/default.html">http://cic.scu.edu.cn/bioinformatics/predict_ppi/default.html</a>	Two fasta sequences
	Prism	<a href="http://cosbi.ku.edu.tr/prism/">http://cosbi.ku.edu.tr/prism/</a>	Two PDB IDs or PDB files
Struct2Net	<a href="http://groups.csail.mit.edu/cb/struct2net/webserver/">http://groups.csail.mit.edu/cb/struct2net/webserver/</a>	Gene names or keywords	

### III. APPLYING NATURAL LANGUAGE PROCESSING TECHNIQUES TO NONCODING RNA IDENTIFICATION

#### 4.1. Relative RNA Prediction Methods.

Arrangement is likewise an essential theme in natural language processing. DNA or RNA arrangements can likewise be seen as content. Succession based numerous grouping arrangement strategies can be utilized just at the succession closeness level. The auxiliary structures of ncRNAs are generally more preserved than their groupings; for instance, miRNA forerunners share the regular clasp like structure and tRNAs frame cloverleaf structures. The elements of numerous ncRNAs are in this manner dictated by their optional structure instead of by their successions. Thus, structure-based different grouping arrangement techniques have been produced to adjust an info succession to known ncRNA structures to decide the ncRNA class to which the information arrangement has a place.

LocARNA can deliver quick and great pairwise and different arrangements of RNA successions. It utilizes a mind boggling RNA vitality show for synchronous collapsing and grouping/structure arrangement of the RNAs. LocARNA performs worldwide and neighborhood succession arrangements and additionally nearby basic arrangement of

RNA atoms. An updated variant of LocARNA, called LocARNA-P, has been created as of late. The new form consolidates a probabilistic model that can process precise various arrangements in view of a probabilistic consistency change and unwavering quality profiles for surveying neighborhood arrangement quality and limiting RNA themes. These elements depend on figuring arrangement and structure coordinate probabilities in light of the LocARNA arrangement demonstrate.

ncRNA classes yet can't recognize individuals from novel ncRNA classes. Most lncRNAs (long noncoding RNAs) can't be anticipated utilizing relative techniques since they don't have particular structures or arrangement comparability. These restrictions imply that similar techniques show low specificity for distinguishing ncRNAs. The different succession arrangement apparatuses that are as of now accessible are recorded in Table 3.

Table 3. Various classification placement tools

Tool	Alignment method	URL
BLAT	Sequence-based	<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>
BLAST		<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
BWA-SW		<a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a>
Multiflign	Structure-based	<a href="http://rma.urmc.rochester.edu/">http://rma.urmc.rochester.edu/</a>
FoldalignM		<a href="http://foldalign.ku.dk/">http://foldalign.ku.dk/</a>
LocARNA/LocARNA-P		<a href="http://www.bioinf.uni-freiburg.de/Software/LocARNA/">http://www.bioinf.uni-freiburg.de/Software/LocARNA/</a>
MASTR		<a href="http://mastr.binf.ku.dk/">http://mastr.binf.ku.dk/</a>
RAF		<a href="http://contra.stanford.edu/contrafold/">http://contra.stanford.edu/contrafold/</a>
RNASampler		<a href="http://tural.wustl.edu/software.html">http://tural.wustl.edu/software.html</a>
RNAshapes		<a href="http://bibiserv.techfak.uni-bielefeld.de/rnashapes/">http://bibiserv.techfak.uni-bielefeld.de/rnashapes/</a>
RNAalifold		<a href="http://www.tbi.univie.ac.at/RNA/">http://www.tbi.univie.ac.at/RNA/</a>
StemLoc		N.A.
MAFFT		<a href="http://mafft.cbrc.jp/alignment/software/index.html">http://mafft.cbrc.jp/alignment/software/index.html</a>
MiRAlign	<a href="http://bioinfo.au.tsinghua.edu.cn/miralign/">http://bioinfo.au.tsinghua.edu.cn/miralign/</a>	

Table 4. miRNA identification procedures

Method	URL	Online service	Local service
MiPred	<a href="http://www.bioinf.seu.edu.cn/miRNA/">http://www.bioinf.seu.edu.cn/miRNA/</a>	✓	✓
microPred	<a href="http://www.cs.ox.ac.uk/people/manohara.rukshan.batuwita/microPred.htm">http://www.cs.ox.ac.uk/people/manohara.rukshan.batuwita/microPred.htm</a>	✓	✓
TripletSVM	<a href="http://bioinfo.au.tsinghua.edu.cn/mirnasvm">http://bioinfo.au.tsinghua.edu.cn/mirnasvm</a>	✓	✓
PlantMiRNAPred	<a href="http://ncab.hit.edu.cn/PlantMiRNAPred/">http://ncab.hit.edu.cn/PlantMiRNAPred/</a>	✓	✓
miRNApre	<a href="http://121.192.180.205:8080/miRNApreWeb/">http://121.192.180.205:8080/miRNApreWeb/</a>	✓	✓
MiReNA	<a href="http://www.ihes.fr/~carbone/data8/">http://www.ihes.fr/~carbone/data8/</a>	✓	✓
HuntMi	<a href="http://adaa.polsl.pl/agudys/huntmi/huntmi.htm">http://adaa.polsl.pl/agudys/huntmi/huntmi.htm</a>	✓	✓
Mirident	<a href="http://www.regulatoryrna.org/pub/mirident">http://www.regulatoryrna.org/pub/mirident</a>	✓	✓
CSHMM	<a href="http://web.iitd.ac.in/~smeet/mirna/">http://web.iitd.ac.in/~smeet/mirna/</a>	✓	✓
HeteroMirPred	<a href="http://ncrna-pred.com/premiRNA.html">http://ncrna-pred.com/premiRNA.html</a>	✓	✓

#### 4.2. Non similar RNA Prediction Methods.

The non-similar strategies are free of homologous data and can, thusly, identify non preserved ncRNAs. Most non relative strategies utilize machine learning systems to make the forecasts, which are like the content mining methods.

As a result of the significance of RNA structure, a few computational RNA collapsing instruments have been created, for example, mfold, RNAfold, vsfold, evofold, and sfold. By and large, these calculations decide the collapsed auxiliary structure from and input arrangement by advancing the intermolecular base matching to limit the free vitality. Some miRNA recognizable proof techniques are appeared in



Table 4 and existing RNA optional expectation devices are recorded in Table 5.

Table 5. Minor estimate tools.

Tool	URL
RNAfold	<a href="http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi">http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi</a>
RNAstructure	<a href="http://rna.urmc.rochester.edu/rnastructure.html">http://rna.urmc.rochester.edu/rnastructure.html</a>
mfold	<a href="http://www.bioinfo.rpi.edu/applications/mfold/">http://www.bioinfo.rpi.edu/applications/mfold/</a>
vsfold	<a href="http://www.rna.it-chiba.ac.jp/~vsfold/vsfold4/">http://www.rna.it-chiba.ac.jp/~vsfold/vsfold4/</a>
evofold	<a href="http://users.soe.ucsc.edu/~jsp/EvoFold/">http://users.soe.ucsc.edu/~jsp/EvoFold/</a>
sfold	<a href="http://sfold.wadsworth.org/cgi-bin/index.pl">http://sfold.wadsworth.org/cgi-bin/index.pl</a>

#### IV. CONCLUSION AND FUTURE RESEARCH

As research on natural language and content mining techniques creates, distinctive application fields will be the way to future reviews. Entomb disciplines spoke to by bioinformatics are turning into the concentration of an expanding number of data science scientists. The use of content mining advances and strategies in bioinformatics study will turn into the concentration of content mining scientists. In the meantime, bioinformatics specialists need to learn content mining advances seriously to take care of particular bioinformatics issues.

In recovering natural writing, aside from the previously mentioned forecast of protein-protein associations and quality malady relationship, numerous issues, especially those that require upgrading writing recovery results, for example, the connections between antagonistic medication response and particle synthesis and also among single nucleotide polymorphism locales, sicknesses, and unfavourable medication impacts, require the utilization of content digging to scan for related learning in a writing database.

In bioinformatics, about all reviews identified with proteomics and anticipating protein structure as indicated by amino corrosive arrangements can be led utilizing content mining and natural language processing innovation. Many develop writings mining innovations, for example, word recurrence insights, condition arbitrary fields, HMM, and setting free linguistic use, have been effectively connected to anticipate optional protein structures, unpredictable districts, associations, and collaboration destinations. Be that as it may, the most recent research brings about content mining and natural language processing ought to be confirmed by applying them in protein and DNA languages. No successful calculation technique is accessible yet to predict third and fourth protein structures, protein homology remote location, protein cluttered locale identification, collaboration arrange foundation, and medication target forecast. Data science analysts ought to create and give more powerful calculations. Moreover, new machine learning and content mining techniques (e.g., semi administered learning and dynamic learning) have been proposed and will be connected in organic writing recovery and bioinformatics. At present, prescribing frameworks in light of input has turned into

another problem area issue in recovering organic writing. What's more, the Hadoop procedure for huge information is another problem area for science groupings.

The improvement of bioinformatics depends on data science. Specifically, content mining and natural language processing specialists ought to give a more broad application space. Specialists of content mining calculations ought to grow more compelling canny calculations in view of the attributes of natural information. This review does not just abridge content mining techniques utilized as a part of bioinformatics and comparing issues, however it likewise gives related sites of fruitful forecast programming. As of late, content mining analysts who are included in bioinformatics can test and look at changed sorts of programming. The creators trust that the quantity of content mining analysts who can apply their own strategies in bioinformatics will expand, which will encourage the improvement of bioinformatics and even hereditary reviews.

#### REFERENCES

- [1] G. Leroy & Hsinchun Chen, "Meeting medical terminology needs-the ontology-enhanced Medical Concept Mapper", in Information Technology in Biomedicine, Vol: 5, Issue: 4, Dec. 2001.
- [2] Cristina Soguero-Ruiz, Luis Lechuga-Suárez, Inmaculada Mora-Jiménez & Javier Ramos-López, "Ontology for Heart Rate Turbulence Domain From The Conceptual Model of SNOMED-CT", in Biomedical Engineering, Vol: 60, Issue: 7, July 2013.
- [3] William Hsu, Ricky K. Taira, Suzie El-Saden & Hooshang Kangarloo, "Context-Based Electronic Health Record: Toward Patient Specific Healthcare", in Information Technology in Biomedicine, Vol: 16, Issue: 2, March 2012.
- [4] Elias Iosif & Alexandros Potamianos, "Unsupervised Semantic Similarity Computation between Terms Using Web Documents", Knowledge and Data Engineering, Vol: 22, Issue: 11, Nov. 2010.
- [5] Marco Masseroli, "Management and Analysis of Genomic Functional and Phenotypic Controlled Annotations to Support Biomedical Investigation and Practice", in Information Technology in Biomedicine, Vol: 11, Issue: 4, July 2007.
- [6] Antonio Sanfilippo, Christian Posse, Banu Gopalan, "Combining Hierarchical and Associative Gene Ontology Relations With Textual Evidence in Estimating Gene and Gene Product Similarity", NanoBioscience, Vol: 6, Issue: 1, March 2007.
- [7] S. Philippi & J. Kohler, "Using XML technology for the ontology-based semantic integration of life science databases", Information Technology in Biomedicine, Volume: 8, Issue: 2, June 2004.