www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 4 April, 2013 Page No. 1097-1105

An Efficient Framework for Name Disambiguition In Digital Library

J.Pricilla

Department of Information Technology

Rajalakshmi Engineering College,
Chennai.

pricilajohn@gmail.com

Abstract

In digital library a number of authors may have same names. The authors with same name may belong to different domains. It leads to name ambiguity while searching the books in digital library. To formalize these problems a unified probabilistic framework is proposed. Using this framework the books titles are analyzed to find out the similarity as well as strong relationship between them. The books titles with strong relationship and higher similarity are grouped to the same cluster. This process continues until all the books in the library are clustered. The users can get the results based on the domain name search in addition with author name search. The users obtain the results by specifying author name and the domain. For this purpose the data in the digital library are also partitioned according to the domain of authors. And hence the authors with the same name are determined easily.

Keywords: *Citation matching, Digital library, Information search and retrieval, Name ambiguity, Publicaion.*

1. INTRODUCTION

A digital library is a library in which collections are stored in digital formats and accessible via computers. Digital library provides the starting point of research. The digital content may be stored locally, or accessed remotely via computer network. A digital library is a type of information retrieval system. In digital library, ambiguous author names occur due to the existence of multiple authors with the same name or different name variations for the same person. Hence we propose an approach that can effectively identify and retrieve information from web pages and use the information to disambiguate authors. Data mining is a relatively young and interdisciplinary field of computer science that results in the discovery of new patterns in large data sets. The overall goal of the data mining process is to extract knowledge from an existing data set and transform it into a human-understandable structure for further use.Mostdigital libraries provide asearchinterfacewhichallows resources to be found. These resources are typically deep web resources since they frequently cannot be located by search engine crawlers. Some digital libraries create special pages or sitemaps to allow search engines to find all their resources. Digital libraries frequently use the Open Archives Initiative Protocol for Metadata Harvesting to expose their metadata to other digital libraries, and search engines like Google Scholar.

There are two general strategies for searching a federation in digital libraries:

1. Distributed searching, and
2. Searching previously harvested metadata.

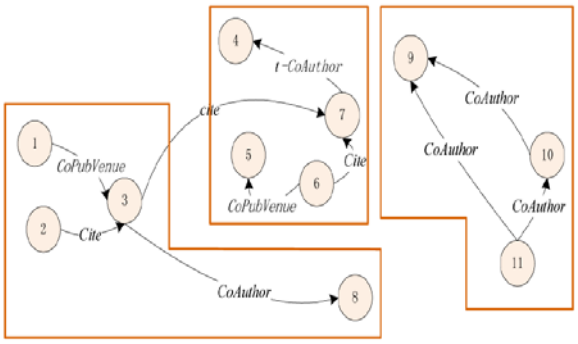
1.1 OVERVIEW

Different people may share identical names in the real World. It is estimated that the 300 most common male Names are used by more than 114 million people in the United States. In many applications name ambiguity will greatly hurt the quality of the retrieved information.
Name ambiguity remains an open problem. Name ambiguity refers to the problem of attributing a publication to a proper

author. This is a common issue in digital library. It is a difficult problem as the same author's name may be written in different ways and different authors may share the same name.

A two step parameter estimation algorithm is proposed to estimate the parameters of disambiguation objective function. Ambiguity is the ability to express more than one interpretation. Name disambiguation mainly fall into three categories: supervised based, unsupervised based, and constraint based. The supervised-based approach tries to learn a specific classification model for each author name from the human-labeled training data. Then, the learned model is used to predict the author assignment of each paper. In the unsupervised based approach clustering algorithms or topic models are employed to find paper Partitions and papers in different partitions are assigned to different authors. The constraint-based approach also Utilizes the clustering algorithms. The difference is that User-provided constraints are used to guide the clustering algorithm toward better data partitioning.

Fig. 1. An example of name disambiguation



Publications and relationships are transformed into an undirected graph, in which each node represents a paper and each edge a relationship. Attributes of a paper are attached to the corresponding node as a feature vector. Name ambiguity can affect the accuracy of citation-based impact analysis and methods to detect ambiguous names are needed. The major tasks of name disambiguation can be defined as the formalizing the disambiguation problem. The formalization needs to consider both local attribute features associated with each paper and relationships between papers and solving the problem in a principled approach. Based on the formalization, propose a principled approach and solve it in an efficient way. This is a nontrivial problem, because most existing clustering methods cannot well balance the two-piece of information. In addition, estimating the number of people is also a challenging task.

Attribute	Description
$p_i.title$	title of p_i
$p_i.pubvenue$	published conference/journal of p_i
$p_i.year$	published year of p_i
$p_i.abstract$	abstract of p_i
$p_i.authors$	authors name set of p_i $\{a_i^{(0)},a_i^{(1)},\ldots a_i^{(u)}\}$
$p_i.references$	references of p_i

TABLE 1
Attributes of Each Publication p_i

A unified framework based on markov random fields author name disambiguation is to detect an ambiguous author name when it is submitted as a query to the citation analysis system. There are basically two types of name ambiguities they are an author may have multiple name variations or multiple authors may share the same name.

TABLE 2
Relationships between Papers

R	W	Relation Name	Description
r_1	w_1	CoPubVenue	$p_i.pubvenue = p_j.pubvenue$
r_2	w_2	CoAuthor	$\exists r, s>0, a_i^{(r)}=a_j^{(s)}$
r_3	w_3	Citation	p_i cites p_j or p_j cites p_i
r_4	w_4	Constraint	feedback supplied by users
r_5	w_5	τ-CoAuthor	τ-extension co-authorship (τ>1)

CoPubVenue ðr1Þ represents two papers published at the same venue. CoAuthor ðr2Þ represents that two papers p1 and p2 have a secondary author with the same name. Citation (r3) represents one paper citing another paper. Constraint (r4) denotes constraints supplied via user. Feedback. Coauthor (r5) represents extension Coauthor relationship. the major tasks of name disambiguation can be defined as: Formalizing the disambiguation problem. The formalization needs to consider both local attribute features associated with each paper and relationships between papers. Solving the problem in a principled approach. Based on the formalization, propose a principled approach and solve it in an efficient way. Determining the number of people K. Given a disambiguation task (without any prior information), determine the actual K. It is nontrivial to perform these tasks. First, it is not immediately clear how to formalize the entire disambiguation problem in a unified framework. Second, some graph models, e.g., Markov Random Field are usually applied to model relational data. However, in the publication informative graph, the papers might be arbitrarily connected by different types of relationships. It is unclear how to perform inference (or parameter

estimation) in such a graph with arbitrary structure. In Addition, estimating the number of people K is also a challenging task.

Name ambiguity can affect the accuracy of citation-based impact analysis and methods to detect ambiguous names are needed. Hence to design an algorithm for the name disambiguation problem by considering both attribute information of the node and the relationships between nodes. Disambiguation is the process of resolving conflicts. In this paper, we propose a unified framework based on Markov Random Fields author name disambiguation is to detect an ambiguous author name when it is submitted as a query to the citation analysis system.

There are basically two types of name ambiguities they are an author may have multiple name variations or multiple authors may share the same name. Name ambiguity can affect the accuracy of citation-based impact analysis and methods to detect ambiguous names are needed. Based on similarity and relationship the papers are clustered. The authors with same name are estimated using bayesian information criteria and two step parameter estimation algorithms. The estimated number of authors with name is very close to original numbers.

In the proposed approach the authors with the same name are displayed at top after we entered the query (that contains the author name as keyword).The users can select the

author name according to their wish and access the books of authors. But in normal search engine the results are not refined. It will be implemented in offline mode. A search platform is used to search the contents of library. Based upon entering the query. Authors with same name are displayed at the top.

1.2 WEAKNESS VS STTRENGHTS

Among the different name disambiguation techniques may not be the most reliable and efficient but it has several advantages over the others:

Digital libraries can be accessed from anywhere around the world. Digital libraries are typically less expensive than traditional libraries. Multiple users can access the same resource at the same time Users can have immediate access to resources access isn't delayed or prohibited due to holds, restrictions for in-library use only, or incorrect shelving, etc.There is no physical degradation of the resource due to handling, storage, or vandalism. Digital libraries aren't limited in terms of size. It can potentially cater better to the needs of users by providing materials that users want and actually use.

Despite the successes of many systems, many issues remain to be addressed. Among those issues, the following are prominent for most systems: The vast information that a digital library can provide can end up becoming a handicap: With the much larger volume of digital information, finding the right material for a specific task becomes increasingly difficult. Digitization violates the copy right law as the thought content of one author can be freely transfer by other without his acknowledgement.

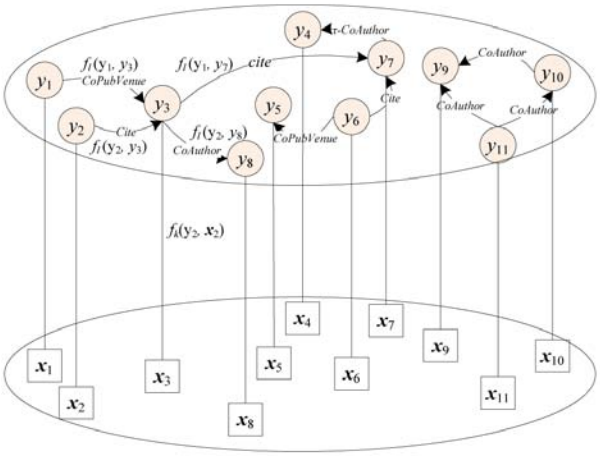
LITERATURE SURVEY

Bailliel, Leif Azzopardi2, and Ian Ruthven[3] proposed a new set of metrics based on the level of assessment which can be used to provide an indication of uncertainty during system comparisons.Comparisons can be detected and investigated easily. The popular systems require more consideration during evaluation.Martin Ester, Rong Ge, Byron J. Gao, Zengjian Hu,Boaz Ben-Moshe[5] implemented on attribute data and relationshipdata.Attributedata describesintrinsiccharacteristics of entities. Relationship data represent extrinsic influences among entities. Connected k-center problem provides the theoretical analysis and degree of relationship. It also provides the criteria relationship between the documents.It improves availability.It reduces quality improvement.Peter T.Davis, David K, Judith L. Klavans[6] focused on the named entity tool. It is used to identify references to a single art object with high precision. It proposes methods to disambiguate intermediate results. It uses the named entities to identify the meaningful segments. It develops the technique into an automatic tool for managing the heterogeneous texts. Accurate results are produced. It is difficult to implement. Indrajit Bhattacharya, Lise Getoor[6] focused on entity resolution is a critical component of data integration. It identifies and consolidates pairs of records or references within the same information sources that are duplicates of each other. It proposes a two-stage collective resolution strategy for processing queries. It extracts and resolves the database references for resolving the query. Queries can be answered efficiently. It works for a database as a whole and not for a specific query.Irvine, CA.DmitriV.Kalashnikov a, Irvine, CA. Sharad Mehrotra Irvine [1] proposed the entity resolution on many different applications. It employs similarity functions that compare values of entity features to determine object descriptions. It can detect the author duplication. The goal of this paper is to group all the entity descriptions that refer to the same real world entities.It is domain independent. It is extensive.Hui Han, Lee Giles,H.Zha,[10] investigated two supervised learning

approaches to disambiguate authors in the citations .One approach uses the naive bayes probability model, a generative model. The other uses Support Vector Machines (SVMs). Based on the parameter estimates, it uses the bayes rule to calculate the probability that each name entry. Improves accuracy, flexibility. It affects the performance of document retrieval.Hui Han,H.Zha, and C.L Giles[7]investigated that an author may have multiple names and multiple authors may share the same name due to name abbreviations. This paper proposes unsupervised learning approach using *K*-way spectral clustering that disambiguates authors in citations. Clustering methods uses cliques to represent a group of names that refer to the same person in his name equivalence identification work. Duplication can be easily avoided. It affects the performance of information retrieval. S.Basu,M.Bilenko[2]focused on the unsupervised clustering. It proposes a probabilistic model that provides a principled framework for incorporating supervision into prototype-based clustering. Semi supervised clustering that employs hidden random markov fields to utilize both labeled and unlabeled data in the clustering process.Improves the clustering quality.It is not adaptable. Ron Bekkerman ,Andrew McCallum[4] proposed two unsupervised frameworks for solving problem such as link structure of the web pages, multi-way distributional clustering method. Social network is tens of times larger than that of grandparents and it grow more with time. The person's homepage may be old and abandoned, containing out of date information, and this may be discovered if it has a broader view on the person's web appearances.Performance can be improved.It is failed to produce the disambiguous person names.Chris Buckley, Ellen M. Voorhees**,** [10] examined how robust the evaluation methodology is to more gross violations of the completeness assumption. It analyzes the effect of imperfect judgment sets by comparing system rankings and repeating the comparisons. It also compares the behavior of the different evaluation measures when complete relevance judgments. It investigates the effect of imperfect judgment. Non relevant document cannot be retrieved.

RESEARCH ELABORATIONS

A Markov Random Field is a conditional probability distribution of labels (hidden variables) that obeys the Markov property. Many special cases of MRF can be developed. A Hidden Markov Random Fields is a member of the family of MRFs and its concept is derived from Hidden Markov Models. two basic observations for the name disambiguation problem. papers with similar content tend to have the same label (belonging to the same author). And papers having strong relationship tend to have the same labels, for example, two papers with coauthors who also author many other papers. An ideal solution is to disambiguate the papers by leveraging both content similarity and paper relationships. This is a nontrivial problem, because most existing clustering methods cannot well balance the two pieces of information.In this paper, we propose a unified framework based on Markov Random Fields.



Graphical representation of the HMRF model. f(yj,yi) and f(xi,xj) are edge feature and node feature, respectively.

A Hidden Markov Random Fields is a member of the family of MRFs and its concept is derived from Hidden Markov Models (HMM).A HMRF is mainly composed of three components: an observable set of random variables, a hidden field of random variables, and neighborhoods between each pair of variables in the hidden field. Formalizing the disambiguation problem as that of grouping relational papers into different clusters. Let the hidden variables be the cluster labels on the papers. Every hidden variable takes a value from the set which are the indexes of the clusters. The observation variables correspond to papers, where every random variable is generated from a conditional probability distribution determined by the corresponding hidden variable Further, the random variables are assumed to be generated conditionally independently from the hidden variables .

Bayesian Information Criterion (BIC) as the criterion to estimate the number of people K. We define an objective function for the disambiguation task. Our goal is to optimize a parameter setting that maximizes the local objective function with some given K and find a number K that maximizes the global objective function. Specifically, we first consider K +1, that is, there is only one person with the given name a. Then, we use a Measurement to determine whether the paper cluster should be split into two sub clusters. Next, for each sub cluster, we again use the measurement to determine whether to split. The operation repeats until some condition is satisfied.

PARAMETER ESTIMATION

The learning algorithm for parameter estimation primarily consists of two iterative steps: Assignment of papers, and Update of parameters The basic idea is that we first randomly choose a Parameter setting and select a centroid for each cluster. Next, we assign each paper to its closest cluster and then calculate the centroid of each paper-cluster based on the assignments. After that, we update the weight of each feature function by maximizing the objective function. For initialization, we randomly assign the value of each parameter .For initialization of the cluster centroid, we first use a graph clustering method to identify the cluster atoms. Basically, papers with similarity less than a threshold will be assigned to disjoint cluster atoms.

PARAMETER ESTIMATION

Algorithm 1. Parameter estimation

Input: $P=\{p_1, p_2, ..., p_n\}$
Output: model parameters Θ and $Y=\{y_1, y_2, ..., y_n\}$, where $y_i \in [1, K]$

1. Initialization
1.1 randomly initialize parameters Θ;
1.2 for each paper x_i, choose an initial value y_i, with $y_i \in [1, K]$;
1.3 calculate each paper cluster centroid $\mu_{(i)}$;
1.4 for each paper x_i and each relationship (x_i, x_j), calculate $f_i(y_i, x_i)$ and $f_k(y_i, y_j)$.
2. Assignment
2.1 assign each paper to its closest cluster centroid;
3. Update
3.1 update of each cluster centroid;
3.2 update of the weight for each feature function.

For initialization of the cluster centroid, we first use a graph clustering method to identify the cluster atoms.Basically; papers with similarity less than a threshold will be assigned to disjoint cluster atoms. We greedily assign papers in the described fashion by always choosing the paper that has the highest similarity to the cluster centroid.

Algorithm 2: One-step sampling
Input: current observation x^0 and labels y^0
Output: sampling results of y^1 and x^1

1: Draw an observation x, from the distribution of $q^0(x_i)$ ($q(x)$ can be obtained by summing over all possible labels);
2: Compute $P(y_i|x)$, the posterior probability distribution over the label variable given the observation x;
3: Compute $P(y_i|y_{\neg i})$, the probability distribution over the label variable given labels of its neighboring observations;
4: Draw a new label y^1_i for each observation from the probability distribution $P(y_i|x)P(y_i|y_{\neg i})$;
5: Given the chosen label, compute the conditional distribution of $P(x_i|y_i)$;
6: Draw each feature of the new observation x^1_i from the conditional distribution $P(x_i|y_i)$.

The similarity function can be easily extended by using any kernel function (e.g., the radius kernel function),

Benefiting from the fact that there are only pairs of papers and pairs of paper-cluster centroid in our
Objective function. With a kernel function, each paper is actually mapped into another new space, which may help disambiguate the papers in some special applications. We tried a few kernel functions, e.g., sigmoid kernel and radius kernel. However, they are not very helpful in our current task. Now, the task is to calculate all parametric terms. The first two terms in are a polynomial combination of the similarity functioned the relational similarity Function which can be calculated.

ESTIMATION OF K

Our strategy for estimating K is to start
by setting it as 1 and we then use the BIC score to measure Whether to split the current cluster. The algorithm runs iteratively. In each iteration, we try to split every cluster C into two sub clusters. We calculate a local BIC score of the new sub model M2. If M2>BIC (M1) then we split the cluster. We calculate a global BIC score for the new model. The process continues by determining if it is possible to split further. Finally, the model with the highest global BIC score is chosen.

Algorithm 3. Estimation of K

Input: $P=\{p_1, p_2, \ldots, p_n\}$
Output: K, $Y=\{y_1, y_2, \ldots, y_n\}$, where $y_i \in [1,K]$

1: i=0, K=1, that is to view P as one cluster: $C^{(i)}=\{C_1\}$;
2: do{
3: foreach cluster C in $C^{(i)}$ {
4: find a best two sub-clusters model M_2 for C;
5: if(BIC(M_2)>BIC(M_1))
6: split cluster C into two sub clusters $C^{(i+1)}=\{C_1, C_2\}$;
7: calculate BIC score for the obtained new model;
8: }while(existing split);
9: choose the model as output with the highest BIC score;

One difficulty in the algorithm might be how to find the Best two sub cluster models for the cluster C . With Different initialization, the resulting sub clusters might be Different. Fortunately, this problem is alleviated in our Framework, benefiting from the cluster atoms identification. In disambiguation, a cluster can consist of several cluster atoms. To split further, we use the cluster atoms as initializing centroids and thus our algorithm tends to result in stable split results.

PairwiseRecall, and PairwiseF1 score, to evaluate our method and to compare with previous methods. The Pairwise measures are adapted for evaluating disambiguation by considering the number of pairs of papers assigned with the same label. Specifically, for any two papers annotated with the same label by the human

annotator, we call it a correct pair. For two papers with the same label predicted by an approach, but do not have the same label in the human annotated data set, we call it a mistakenly predicted pair.
Thus, we can define the measures as follows:

PairwisePrecision
= # PairsCorrectlyP redictedToSameAuthor
TotalPairsP redictedToSameAuthor

PairwiseRecall
=# PairsCorrectlyP redictedToSameAuthor
Total PairsToSameAuthor

PairwiseF1
= 2 * PairwisePrecision * Pairwise Recall
Pairwise precision +þairwiseRecall

MODULE TITTLE

3.1 Data sets storage

The main idea is to create a database containing all the publications from different domains. The publications from different domains are collected and stored in database. The administrator maintains these resources in database. Then it creates a search platform to search the publications that are available in the database.

3.2 Clustering the applications

The attribute information of publications is calculated and it is known as node features of the corresponding publication. Then find out the relationship measure between publications in the database. Based on similarity and relationship, cluster the publications. Then split the cluster into sub clusters by considering higher similarity between publications.

3.3 Estimation of number of persons with same name

An author name is taken as input by the system to find the number of persons with the same name. It searches all the publications and compares the author name with the attributes of publications. Then find the number of persons for the particular author name. Repeat this procedure for all authors.

3.4 Refining search results

Initially user enters their queries with author name as a keyword (without mentioning domain name). The results shown to them consist of list of authors with same name. By selecting desired author name users can read the publications. Now the users enter their queries with domain name in addition to author name. Now the refined results are shown to the users.

4.1 ARCHITECTURE DIAGRAM

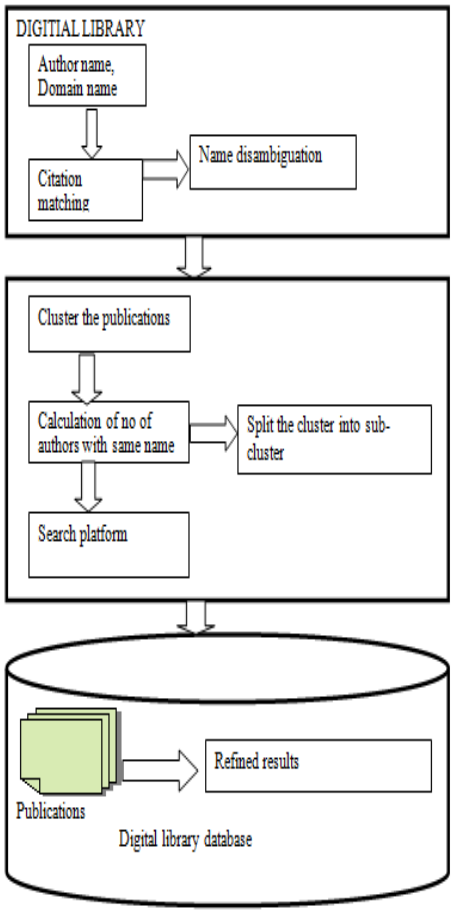


Figure 4.1 Proposed architecture diagram

Figure 4.1 represents the architecture diagram of the system. It is three tire architecture consists of an digital library users, clusteringfuntionalities and database has been connected with business logic tire. In digital library can have author name, domain name is considered as a input to the storage of contents in the digital library database. Once the dataset values are entered it leads to the number of authors with same name are grouped or not. The citation matching is the

process of estimating the same author further improvement in edition, when number of author, number of domain is grouped as cluster. It leads the production of name disambiguation results. The name disambiguation is the one in which it produces the result as the number of authors with more than one meaning. In the next tier the clustering process refers to the grouping of on the book details. By this process it can be able to find the relationship between the book details such as in the case of any relation between the publications venues, author. The relation between the author and the coauthor are estimated. The co citation is the one which matches the co publication venue and the co author relationship. It leads to the calculation of number of authors with the same name. The cluster is then partioned into the sub cluster by grouping some of the book details. The search platform displays the results from the library The normal digital library interact only for specific organization. Search platform it can be useful to search and retrieve the book contents. The third tier is the one which can produces the refined results. The collection of publications can lead to the storage of results in the digital library database.

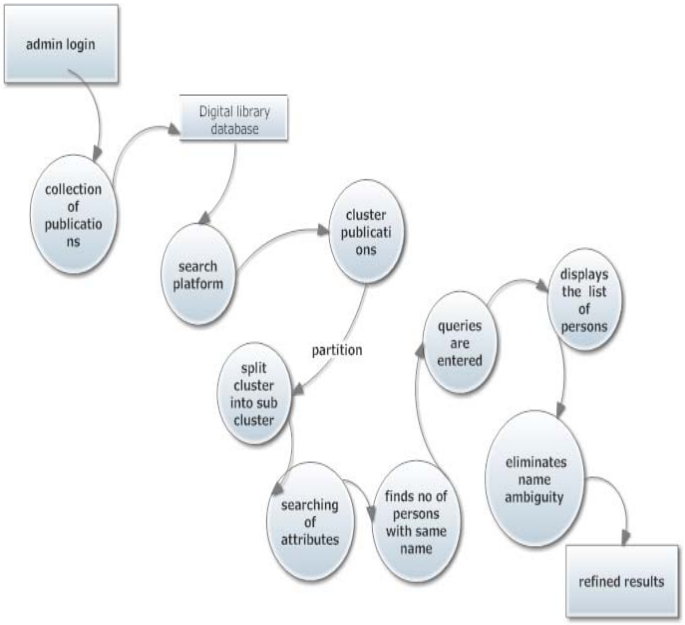


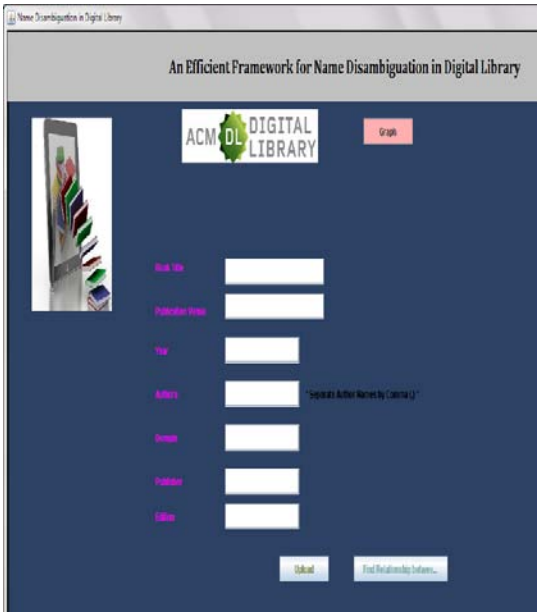
Figure 4.2 Proposed DFD diagram

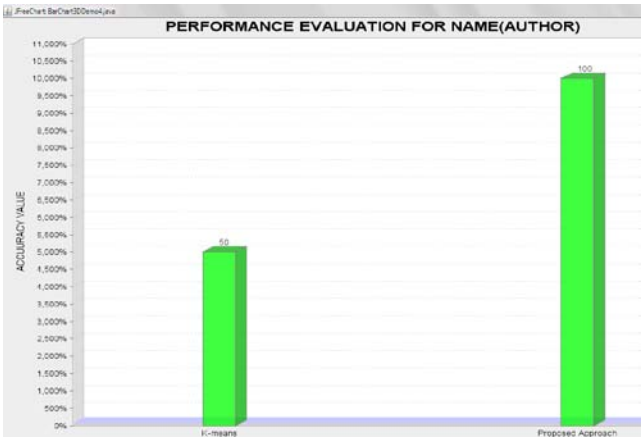
Figure 4.2 represents the overall DFD for the proposed system.The search platform can capture the attribute information of every publication.The attribute information of publications is used to find out the relationshipmeasure between publications in the database. It can determine the number of persons for the particular author name. The result displays the list of authors with same name and it also shows the no of publications. By selecting desired author name users can read the publications. It leads to the elimination of name ambiguity. Now the refined results are shown to the users.

2. CONCLUSION AND FUTURE WORK

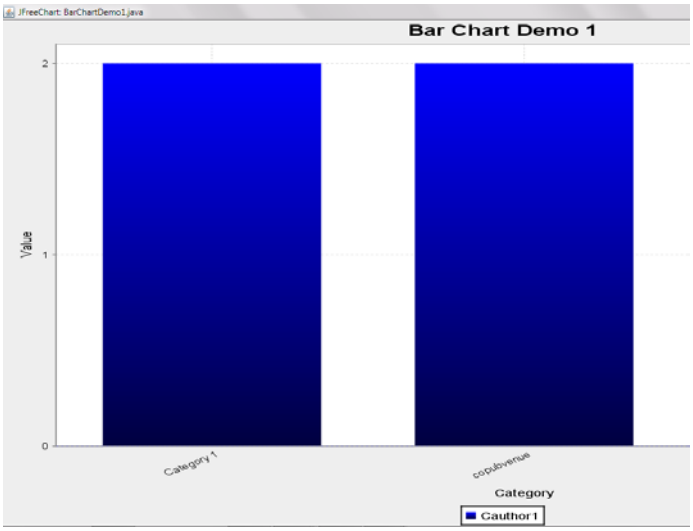
The design specifications for the "An Efficient Framework for Name Disambiguation in Digital Library" have been drafted out. The proposed work is found feasible and is believed that this system can overcome the difficulties prevalent in the existing system. It is intended to provide the refined results while searching the books and to eliminate the name ambiguity in digital library. The analysis on the requirements and a design for the proposed system has been screened. The requirement analysis process includes learning and determining about the working environment, technical requirements and logical aspects or features of the system. The design of the system has been sketched out using this analyzed information.The design has to be implemented in future (Phase 2 of Project Work), this design is applicable to certain changes as and when required in order to develop a prototype for the proposed work. The changes inserted would merely be in the physical components or other dependent components alone. The logical design of the system and its functionality would be preserved.

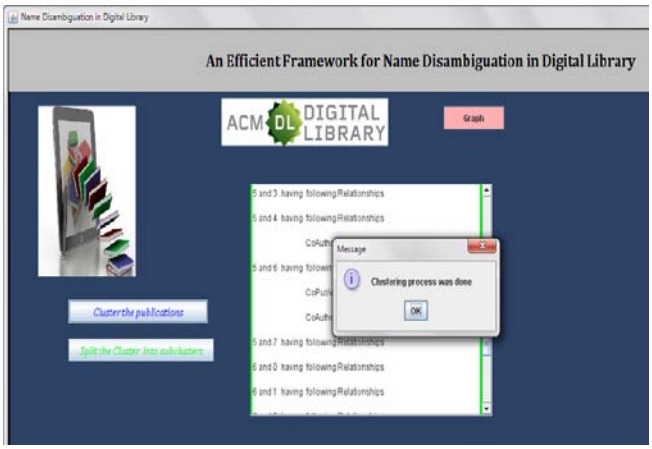
SCREEN SHOTS

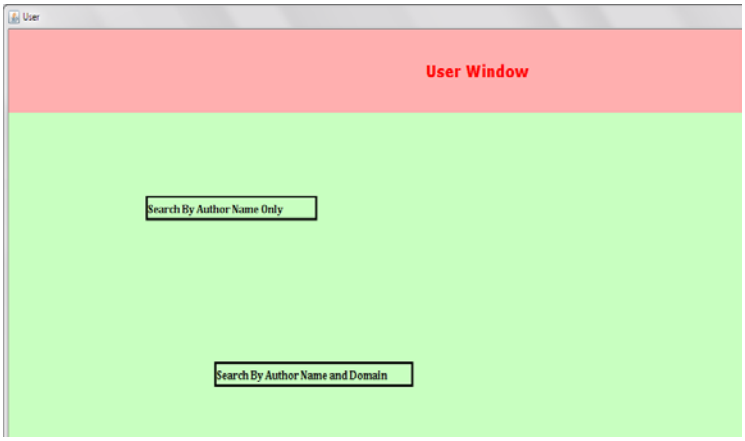


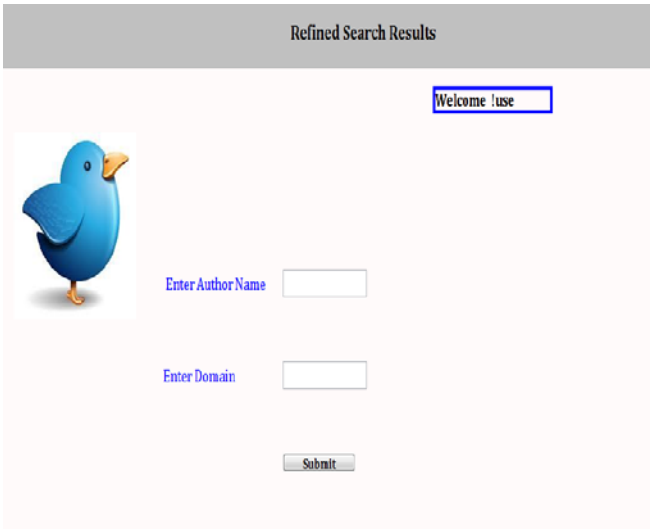

PERFORMANCE EVALUATION FOR NAME(AUTHOR)
ACCURACY VALUE
50
100
K-means
Proposed Approach

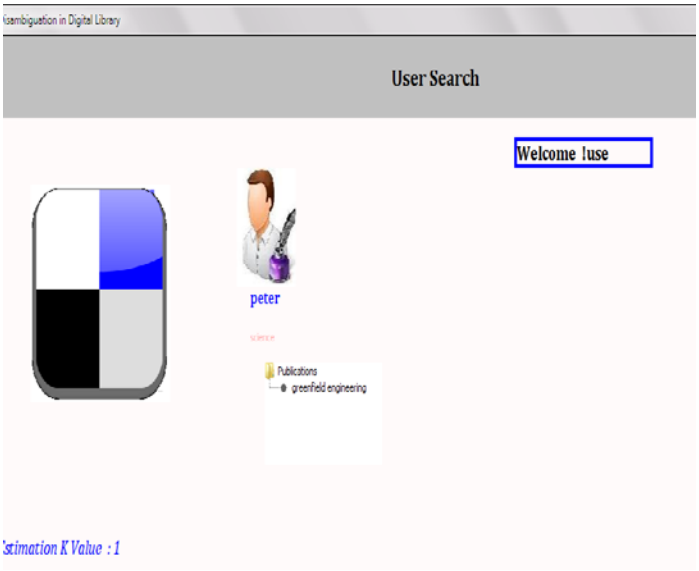

Name Disambiguation in Digital Library
An Efficient Framework for Name Disambiguation in Digital Library
ACM DL DIGITAL LIBRARY
Graph
5 and 3 having following Relationships
5 and 4 having following Relationships
CoAuthor
5 and 6 having following Relationships
CoPubVenue
CoAuthor
5 and 7 having following Relationships
8 and 0 having following Relationships
8 and 1 having following Relationships
Cluster the publications
Split the Cluster into subclusters


JFreeChart: BarChartDemo1.java
Bar Chart Demo 1
Value
Category 1
copubvenue
Category
Cauthor1


Name Disambiguation in Digital Library
An Efficient Framework for Name Disambiguation in Digital Library
ACM DL DIGITAL LIBRARY
Graph
Message
Clustering process was done
OK
Cluster the publications
Split the Cluster into subclusters







REFERENCES

[1] Zhaoqi Che, Dmitri, V, Kalashnikov, and S.Mehtrotra "Adaptive Graphical Approach to Entity Resolution".

[2]S.Basu, M.Bilenko, and R.I Mooney,"Probabilistic Framework for Semi Supervised Clustering".

[3]C.Buckley and E.M Voorhees,"Retrieval Evaluation Methodology for Incomplete Relevance Assessments" Mark Baillie1, Leif Azzopardi2, and Ian Ruthven11 Department of Computing and Information Sciences, University of Strathclyde, Glasgow, Department ofComputingScience, University of Glasgow, Glasgow, UK.

[4Andrew,McCallum," Disambiguating Web Appearances of People in a Social Network" ,Bekkerman Dept. of Computer Science University of Massachusetts Amherst, MA 01003, USA, Andrew McCallum Dept. of Computer Science, University of Massachusetts, Amherst, USA.

[5] Martin Ester,RongGe,Byron,"Joint Cluster Analysis of Attribute Data and Relationship Data: the Connected k-Center Problem",Ben-MosheSchool of Computing Science, Simon Fraser University.

[6] David K. Elson, Peter T.Davis ,"Methods for Precise Named Entity Matching in Digital Collections,"Columbia University, New York, NY 10027, David K. Elson Columbia University, New York, NY 10027, Judith L. Klavans Columbia University, NewYork, NY 10027.

[7]Hui Han,C.L Giles,H.Zha"Name Disambiguation in Author Citations using a K-way Spectral Clustering Method" First Avenue Sunnyvale, CA, Department of Computer Science and Engineering The Pennsylvania State University , PA, 16802.

[8]Indrajit,Lise Getoor,"Online Collective Entity Resolution" IBM India Research Lab New Delhi, India, Computer Science Dept. University of Maryland, College Park.

[9]Chris Buckley Sabir Research, Inc. Gaithersburg, MD 20878, EllenM.Voorhees ,"Retrieval Evaluation with Incomplete Information"National Institute of Standards and Technology, Gaithersburg, Maryland 20899.