

INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT & MANAGEMENT

I
J
R
C
M



A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.

as well as in Open J-Gate, India [link of the same is duly available at infibnet of University Grants Commission (U.G.C.)]

Registered & Listed at: Index Copernicus Publishers Panel, Poland & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 1667 Cities in 145 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

www.ijrcm.org.in

CONTENTS

Sr. No.	TITLE & NAME OF THE AUTHOR (S)	Page No.
1.	EFFICIENCY AND PERFORMANCE OF e-LEARNING PROJECTS IN INDIA SANGITA RAWAL, DR. SEEMA SHARMA & DR. U. S. PANDEY	1
2.	AN ADAPTIVE DECISION SUPPORT SYSTEM FOR PRODUCTION PLANNING: A CASE OF CD REPLICATOR SIMA SEDIGHADELI & REZA KACHOUIE	5
3.	CONSTRUCT THE TOURISM INTENTION MODEL OF CHINA TRAVELERS IN TAIWAN WEN-GOANG, YANG, CHIN-HSIANG, TSAI, JUI-YING HUNG, SU-SHIANG, LEE & HUI-HUI, LEE	9
4.	FINANCIAL PLANNING CHALLENGES AFFECTING IMPLEMENTATION OF THE ECONOMIC STIMULUS PROGRAMME IN EMBU COUNTY, KENYA PAUL NJOROGI THIGA, JUSTO MASINDE SIMIYU, ADOLPHUS WAGALA, NEBAT GALO MUGENDA & LEWIS KINYUA KATHUNI	15
5.	IMPACT OF ELECTRONIC COMMERCE PRACTICES ON CUSTOMER E-LOYALTY: A CASE STUDY OF PAKISTAN TAUSIF M. & RIAZ AHMAD	22
6.	SOCIAL NETWORKING IN VIRTUAL COMMUNITY CENTRES: USES AND PERCEPTION AMONG SELECTED NIGERIAN STUDENTS DR. SULEIMAN SALAU & NATHANIEL OGUCHE EMMANUEL	26
7.	EXPOSURE TO CLIMATE CHANGE RISKS: CROP INSURANCE DR. VENKATESH. J, DR. SEKAR. S, AARTHY.C & BALASUBRAMANIAN. M	32
8.	SCENARIO OF ENTERPRISE RESOURCE PLANNING IMPLEMENTATION IN SMALL AND MEDIUM SCALE ENTERPRISES DR. G. PANDURANGAN, R. MAGENDIRAN, L.S. SRIDHAR & R. RAJKOKILA	35
9.	BRAIN TUMOR SEGMENTATION USING ALGORITHMIC AND NON ALGORITHMIC APPROACH K.SELVANAYAKI & DR. P. KALUGASALAM	39
10.	EMERGING TRENDS AND OPPORTUNITIES OF GREEN MARKETING AMONG THE CORPORATE WORLD DR. MOHAN KUMAR. R, INITHA RINA.R & PREETHA LEENA .R	45
11.	DIFFUSION OF INNOVATIONS IN THE COLOUR TELEVISION INDUSTRY: A CASE STUDY OF LG INDIA DR. R. SATISH KUMAR, MIHIR DAS & DR. SAMIK SOME	51
12.	TOOLS OF CUSTOMER RELATIONSHIP MANAGEMENT – A GENERAL IDEA T. JOGA CHARY & CH. KARUNAKER	56
13.	LOGISTIC REGRESSION MODEL FOR PREDICTION OF BANKRUPTCY ISMAIL B & ASHWINI KUMARI	58
14.	INCLUSIVE GROWTH: REALITY OR MYTH IN INDIA DR. KALE RACHNA RAMESH	65
15.	A PRACTICAL TOKENIZER FOR PART-OF SPEECH TAGGING OF ENGLISH TEXT BHAIRAB SARMA & BIPUL SHYAM PURKAYASTHA	69
16.	KEY ANTECEDENTS OF FEMALE CONSUMER BUYING BEHAVIOR WITH SPECIAL REFERENCE TO COSMETICS PRODUCT DR. RAJAN	72
17.	MANAGING HUMAN ENCOUNTERS AT CLASSROOMS - A STUDY WITH SPECIAL REFERENCE TO ENGINEERING PROGRAMME, CHENNAI DR. B. PERCY BOSE	77
18.	THE IMPACT OF E-BANKING ON PERFORMANCE – A STUDY OF INDIAN NATIONALISED BANKS MOHD. SALEEM & MINAKSHI GARG	80
19.	UTILIZING FRACTAL STRUCTURES FOR THE INFORMATION ENCRYPTING PROCESS UDAI BHAN TRIVEDI & R C BHARTI	85
20.	IMPACT OF LIBERALISATION ON PRACTICES OF PUBLIC SECTOR BANKS IN INDIA DR. R. K. MOTWANI & SAURABH JAIN	89
21.	THE EFFECTIVENESS OF PERFORMANCE APPRAISAL ON ITES INDUSTRY AND ITS OUTCOME DR. V. SHANTHI & V. AGALYA	92
22.	CUSTOMERS ARE THE KING OF THE MARKET: A PRICING APPROACH BASED ON THEIR OPINION - TARGET COSTING SUSANTA KANRAR & DR. ASHISH KUMAR SANA	97
23.	WHAT DRIVE BSE AND NSE? MOCHI PANKAJKUMAR KANTILAL & DILIP R. VAHONIYA	101
24.	A CASE APPROACH TOWARDS VERTICAL INTEGRATION: DEVELOPING BUYER-SELLER RELATIONSHIPS SWATI GOYAL, SONU DUA & GURPREET KAUR	108
25.	ANALYSIS OF SOURCES OF FRUIT WASTAGES IN COLD STORAGE UNITS IN TAMILNADU ARIVAZHAGAN.R & GEETHA.P	113
26.	A NOVEL CONTRAST ENHANCEMENT METHOD BY ARBITRARILY SHAPED WAVELET TRANSFORM THROUGH HISTOGRAM EQUALIZATION SIBIMOL J	119
27.	SCOURGE OF THE INNOCENTS A. LINDA PRIMLYN	124
28.	BUILDING & TESTING MODEL IN MEASUREMENT OF INTERNAL SERVICE QUALITY IN TANCEM – A GAP ANALYSIS APPROACH DR. S. RAJARAM, V. P. SRIRAM & SHENBAGASURIYAN.R	128
29.	ORGANIZATIONAL CREATIVITY FOR COMPETITIVE EXCELLENCE REKHA K.A	133
30.	A STUDY OF STUDENT'S PERCEPTION FOR SELECTION OF ENGINEERING COLLEGE: A FACTOR ANALYSIS APPROACH SHWETA PANDIT & ASHIMA JOSHI	138
	REQUEST FOR FEEDBACK	146

CHIEF PATRON

PROF. K. K. AGGARWAL

Chancellor, Lingaya's University, Delhi
Founder Vice-Chancellor, Guru Gobind Singh Indraprastha University, Delhi
Ex. Pro Vice-Chancellor, Guru Jambheshwar University, Hisar

FOUNDER PATRON

LATE SH. RAM BHAJAN AGGARWAL

Former State Minister for Home & Tourism, Government of Haryana
Former Vice-President, Dadri Education Society, Charkhi Dadri
Former President, Chinar Syntex Ltd. (Textile Mills), Bhiwani

CO-ORDINATOR

AMITA

Faculty, Government M. S., Mohali

ADVISORS

DR. PRIYA RANJAN TRIVEDI

Chancellor, The Global Open University, Nagaland

PROF. M. S. SENAM RAJU

Director A. C. D., School of Management Studies, I.G.N.O.U., New Delhi

PROF. M. N. SHARMA

Chairman, M.B.A., Haryana College of Technology & Management, Kaithal

PROF. S. L. MAHANDRU

Principal (Retd.), Maharaja Agrasen College, Jagadhri

EDITOR

PROF. R. K. SHARMA

Professor, Bharti Vidyapeeth University Institute of Management & Research, New Delhi

CO-EDITOR

DR. BHAVET

Faculty, M. M. Institute of Management, Maharishi Markandeshwar University, Mullana, Ambala, Haryana

EDITORIAL ADVISORY BOARD

DR. RAJESH MODI

Faculty, Yanbu Industrial College, Kingdom of Saudi Arabia

PROF. SANJIV MITTAL

University School of Management Studies, Guru Gobind Singh I. P. University, Delhi

PROF. ANIL K. SAINI

Chairperson (CRC), Guru Gobind Singh I. P. University, Delhi

DR. SAMBHAVNA

Faculty, I.I.T.M., Delhi

DR. MOHENDER KUMAR GUPTA

Associate Professor, P. J. L. N. Government College, Faridabad

DR. SHIVAKUMAR DEENE

Asst. Professor, Dept. of Commerce, School of Business Studies, Central University of Karnataka, Gulbarga

DR. MOHITA

Faculty, Yamuna Institute of Engineering & Technology, Village Gadholi, P. O. Gadholi, Yamunanagar

ASSOCIATE EDITORS

PROF. NAWAB ALI KHAN

Department of Commerce, Aligarh Muslim University, Aligarh, U.P.

PROF. ABHAY BANSAL

Head, Department of Information Technology, Amity School of Engineering & Technology, Amity University, Noida

PROF. A. SURYANARAYANA

Department of Business Management, Osmania University, Hyderabad

DR. SAMBHAV GARG

Faculty, M. M. Institute of Management, Maharishi Markandeshwar University, Mullana, Ambala, Haryana

PROF. V. SELVAM

SSL, VIT University, Vellore

DR. PARDEEP AHLAWAT

Associate Professor, Institute of Management Studies & Research, Maharshi Dayanand University, Rohtak

DR. S. TABASSUM SULTANA

Associate Professor, Department of Business Management, Matrusri Institute of P.G. Studies, Hyderabad

SURJEET SINGH

Asst. Professor, Department of Computer Science, G. M. N. (P.G.) College, Ambala Cantt.

TECHNICAL ADVISOR

AMITA

Faculty, Government H. S., Mohali

DR. MOHITA

Faculty, Yamuna Institute of Engineering & Technology, Village Gadholi, P. O. Gadholi, Yamunanagar

FINANCIAL ADVISORS

DICKIN GOYAL

Advocate & Tax Adviser, Panchkula

NEENA

Investment Consultant, Chambaghat, Solan, Himachal Pradesh

LEGAL ADVISORS

JITENDER S. CHAHAL

Advocate, Punjab & Haryana High Court, Chandigarh U.T.

CHANDER BHUSHAN SHARMA

Advocate & Consultant, District Courts, Yamunanagar at Jagadhri

SUPERINTENDENT

SURENDER KUMAR POONIA

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email address: infoijrcm@gmail.com.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: SUBMISSION OF MANUSCRIPT IN THE AREA OF _____.

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled ' _____ ' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:

Designation:

Affiliation with full address, contact numbers & Pin Code:

Residential address with Pin Code:

Mobile Number (s):

Landline Number (s):

E-mail Address:

Alternate E-mail Address:

NOTES:

- a) The whole manuscript is required to be in **ONE MS WORD FILE** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
- b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)
- c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
- d) The total size of the file containing the manuscript is required to be below **500 KB**.
- e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
- f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name, designation, affiliation (s), address, mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.
6. **MANUSCRIPT:** Manuscript must be in **BRITISH ENGLISH** prepared on a standard A4 size **PORTRAIT SETTING PAPER**. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.
7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.
9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed **5000 WORDS**.

10. **FIGURES & TABLES:** These should be simple, crystal clear, centered, separately numbered & self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.
11. **EQUATIONS:** These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.
12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:
 - All works cited in the text (including sources for tables and figures) should be listed alphabetically.
 - Use (ed.) for one editor, and (ed.s) for multiple editors.
 - When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
 - Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
 - The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
 - For titles in a language other than English, provide an English translation in parentheses.
 - The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 <http://epw.in/user/viewabstract.jsp>

A PRACTICAL TOKENIZER FOR PART-OF SPEECH TAGGING OF ENGLISH TEXT

BHAIRAB SARMA
RESEARCH SCHOLAR
DEPARTMENT OF COMPUTER SCIENCE
ASSAM UNIVERSITY
SILCHAR

BIPUL SHYAM PURKAYASTHA
PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
ASSAM UNIVERSITY
SILCHAR

ABSTRACT

Tokenization is an important task in part of speech tagging. A token is a tiny part of a sentence with individual meaning. In part-of-speech tagging, all taggers must tokenize each input sentence into smaller parts before classified. The performance of a tagger is based on how accurately its tokenizer tokenizes its given input sentence. Depending on grammatical and inflectional rules, different approaches are used for different languages in taggers. In this paper we present a practical approach of English text tokenization. Although English is comparatively less morphologically inflected language, there are some special issues that should be considering in POS tagging. We develop a process with some special consideration which tokenizes sentences in multiple succession, so that maximum accuracy could be expected. We excluded formatted texts, graphics, tables and images from our consideration. Here a user can upload a text file written in English and the model separates each component into an array and taking each part into special consideration given in the sequence of token that consist in the file. The output of this model can input to the analyser for lexical analysis and tagged with proper tags.

KEYWORDS

inflectional rule, POST, tagger, tokenizer, WSD.

INTRODUCTION

The objective of Natural Language Processing (NLP) is to application of natural languages in computer applications. Under NLP paradigm there are many tasks used for machine translation. Part-of-speech tagging is one of the activities that assign proper symbols to each word from a written text of natural language. A language is composed of sentences which composed of words. Apart from word it may consist of numbers, punctuation marker and grouping word which represent one unique meaning. In statistical tagger¹, tagging is done based on contextual probability of word meaning. To retrieve the proper meaning from the context, it is very much important to tokenize the group of words and its associated symbols. S. Kulick in² explained a successive method for tokenization of Arabic text. In his approach, named entities were not considered separately in terms of token.

In linguistics, there are at least two types of token classes: sentence boundary and words are considered. However in the context of tagging, other classes of token needed to be considered that may include numbers, paragraph boundaries and various sorts of punctuation symbols etc. Aho et al. (1986) specify a sentence boundary with the period followed by a white space, followed by an uppercase letter. For example in *www.rediffmail.com*, although three periods are present in this sample word, only one token should be identified, although there are three word separated by three period but due to the absence of white space. Moreover during tagging the entire line is classed as a single part as Foreign word [FW]. In our approach we tokenize each word through successive tokenization method where in first, we separate each word by a white space. If two or more continuous words begin with capital letter without any separator, we consider them under single token. For example "Veenit Chaitanya, Akshar Bharti & Rajeev Segal" when tokenize with our model, we get three tokens as three different names separated by comma and ampersand. Kupiec (1992) developed a practical tagger using HMM approach using regular expression for tokenizing string³. In [2], using regular expression, a substring is extracted and assign in a new string with an internal name and a list is maintained for core POS tags. This list is returned for assigning tags.

OUR APPROACH

In our approach we tokenize a given sentence in different succession. The processes adopted here are explained below. Our model is enabling to take a sentence input directly from user or can upload a file from hard disk of the client machine. Here the type of the uploaded file should be in text format either .txt or .doc. If the file is processed in .docx format, before uploading it should be manually saved as .doc format. We consider here only simple text, not the formatted text or any other style like word art or rich document. A document consists of tables, images and graphics will show garbage. The task of formatted text processing is left for future research. The process sequence is given figure 1 below:

FIGURE 1: PROCESS OUTLINE



EXPERIMENT

We experimenting our approach in various ways. A simple sentence is uploaded directly from the client machine and tokenize successively for words, separated by white spaces. Next we look through for word having punctuation marker (period, comma, question marker). We used a temporary string to store the component string and consider for further analysis. We achieve 90% accuracy for plain text. Rest 10% errors for ambiguity with named entity and abbreviation word present in sentences. For example, in the sentence: "Mr. Ravishankar Nair is a man of moral". During tokenization through our model, we the get the output as [Mr][.][Ravishankar Nair][is][a][man][of][moral] due to ambiguity with period after Mr. However for the name, we are enabled to properly identify its expandability. After properly tagging, we get following output (using Penn treebank tagset).

FIGURE 2: A SAMPLE OUTPUT OF THE MODEL

Output without Correction	Corrected Output
[Mr]/[FW]	[Mr. Ravishankar Nair]/[NNP]
[.]/[.]	[is]/[MD]
[Ravishankar Nair]/[NNP]	[a]/[DT]
[is]/[MD]	[man]/[NN]
[a]/[DT]	[of]/[IN]
[man]/[NN]	[moral]/[JJ]
[of]/[IN]	
[moral]/[JJ]	

Here the achieve accuracy level is approximately 87.5%. Now we test with another sentence. Let the input sentence to be tokenize is "LIFER by Hendrix (1978) and INTELLECT by Harris (1977) were some of the early systems of automated machine translation. (Akshar Bharati, 1985)"

In the given sentence if we tokenize each word separated by while space the output array will be

0	1	2	3	4	5	6	7	8
LIFER	by	Hendrix	(1977)	and	INTELLECT	by	Harris	(1977)
9	10	11	12	13	14	15	16	17
were	some	of	the	early	systems	of	automated	machine
18	19	20	21					
translation.	(Akshar	Bharti,	1985)					

In the sample array the difficulties are, in token 3(1987), 8(1977), 19, 20 & 21 due to the present of parenthesis. Here the two parentheses are individual token falls under punctuation. From the POST point of view, token 19 & 20 should be assign by one tag because these two word (Akshar and Bharti) represent one name, category proper noun, singular number and assign tag is [NNP]. However in first iteration, 'Akshar' and 'Bharti' comes in two separate token. Secondly, in token 18 and 20, two punctuation mark period '.' and comma ',' comes together in a single token. In our approach we check each case by case separately and tokenize accordingly. To examine our model online, user can test by link (http://www.apostane/data_input.php).

ALGORITHM USED

Input: S - the input string to be tokenize

F - a temporary file created on disk

A- Array of words

Output: Q- sequence of tags

1. Upload the file to the server
2. Read the strings from the file and save into temporary file
3. Separate the words with white spaces
4. While not eof()
5. W= Reads a word from the list and check each character (C) in W
6. If C is punctuation marker
7. Split the word in two classes
8. A=splitted word 1
9. B=punctuation symbol
10. Store A & B into the list
11. If List[i] & List[i+1] have begin with Capital Mark // for name entity recognition

12. Marge(List[i],List[i+1]) and stored in list
13. End while
14. Return List(Q)

Now the output of this model is input to the morphological analyzer where the inflectional information of the word is extracted. The root word is steamed to categorize the word for tagging. Now the new list to be tagged is the list of root word coming after morphological analyzer.

CONCLUSION

We test our model with different text and uploaded file in text format and we achieve 87.5% accuracy with plain text irrespective morphological inflection. Presently we are applying this model for word sense disambiguation, separating nouns with other part-of-speech and punctuation mark identification. This approach can also be used to separate inflectional component of an inflected word with little modification. This task we left for future work during morphological analysis. Tokenization does not consider the inflected word separately. In this model we do not pay more attention for tokenize for local word grouping. For LWG, separate algorithms have to use based on sense of thee words or more using Markov Model.

REFERENCES

1. Scot M. Thede and Mary P Harper, A Second order Hidden Morkov Model for Part-of-Speech Tagging, In Proceeding of the 37th Annual Meeting of the ACL
2. Seth Kulick, Simultaneous Tokenization and Part-of-Speech Tagging for Arabic without a Morphological Analyzer, Proceedings of the ACL 2010 Conference Short Papers, pages 342–347, Uppsala, Sweden, 11-16 July 2010. c 2010 Association for Computational Linguistics
3. Doug Cutting, Julian Kupiec, Jan Pedersen and Penelope Sibun, A Practical Part-of-Speech Tagger, Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Commerce, IT and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail i.e. **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail infoijrcm@gmail.com.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals

