Vol 5 Issue 10 July 2016

ISSN No : 2249-894X

Monthly Multidisciplinary Research Journal

Review Of Research Journal

Chief Editors

Ashok Yakkaldevi
A R Burla College, India

Ecaterina Patrascu
Spiru Haret University, Bucharest

Kamani Perera
Regional Centre For Strategic Studies,
Sri Lanka

Welcome to Review Of Research

RNI MAHMUL/2011/38595 **ISSN No.2249-894X**

Review Of Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial Board readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Regional Editor

Manichander Thammishetty
Ph.d Research Scholar, Faculty of Education IASE, Osmania University, Hyderabad.

Advisory Board

Kamani Perera
Regional Centre For Strategic Studies, Sri Lanka

Ecaterina Patrascu
Spiru Haret University, Bucharest

Fabricio Moraes de AlmeidaFederal
University of Rondonia, Brazil

Anna Maria Constantinovici
AL. I. Cuza University, Romania

Romona Mihaila
Spiru Haret University, Romania

Delia Serbescu
Spiru Haret University, Bucharest, Romania

Xiaohua Yang
University of San Francisco, San Francisco

Karina Xavier
Massachusetts Institute of Technology (MIT), USA

May Hongmei Gao
Kennesaw State University, USA

Marc Fetscherin
Rollins College, USA

Liu Chen
Beijing Foreign Studies University, China

Mabel Miao
Center for China and Globalization, China

Ruth Wolf
University Walla, Israel

Jie Hao
University of Sydney, Australia

Pei-Shan Kao Andrea
University of Essex, United Kingdom

Loredana Bosca
Spiru Haret University, Romania

Ilie Pintea
Spiru Haret University, Romania

Mahdi Moharrampour
Islamic Azad University buinzahra Branch, Qazvin, Iran

Titus Pop
PhD, Partium Christian University, Oradea,
Romania

J. K. VIJAYAKUMAR
King Abdullah University of Science & Technology,Saudi Arabia.

George - Calin SERITAN
Postdoctoral Researcher
Faculty of Philosophy and Socio-Political Sciences
Al. I. Cuza University, Iasi

REZA KAFIPOUR
Shiraz University of Medical Sciences
Shiraz, Iran

Rajendra Shendge
Director, B.C.U.D. Solapur University, Solapur

Nimita Khanna
Director, Isara Institute of Management, New Delhi

Salve R. N.
Department of Sociology, Shivaji University, Kolhapur

P. Malyadri
Government Degree College, Tandur, A.P.

S. D. Sindkhedkar
PSGVP Mandal's Arts, Science and Commerce College, Shahada [M.S.]

Anurag Misra
DBS College, Kanpur

C. D. Balaji
Panimalar Engineering College, Chennai

Bhavana vivek patole
PhD, Elphinstone college mumbai-32

Awadhesh Kumar Shirotriya
Secretary, Play India Play (Trust),Meerut (U.P.)

Govind P. Shinde
Bharati Vidyapeeth School of Distance Education Center, Navi Mumbai

Sonal Singh
Vikram University, Ujjain

Jayashree Patil-Dake
MBA Department of Badruka College Commerce and Arts Post Graduate Centre (BCCAPGC),Kachiguda, Hyderabad

Maj. Dr. S. Bakhtiar Choudhary
Director,Hyderabad AP India.

AR. SARAVANAKUMARALAGAPPA UNIVERSITY, KARAIKUDI,TN

V.MAHALAKSHMI
Dean, Panimalar Engineering College

S.KANNAN
Ph.D , Annamalai University

Kanwar Dinesh Singh
Dept.English, Government Postgraduate College , solan

More.....

ISSN: 2249-894X Impact Factor : 3.1402(UIF) Volume - 5 | Issue - 10 | July - 2016

Review Of Research

हिंदी पाठ भाव–विश्लेषण हेतु संगणकीय मॉडल

प्रितेंद्र कुमार मालाकार , प्रवेश कुमार द्विवेदी
प्रौद्योगिकी अध्ययन केंद्र (भाषा विद्यापीठ) म. गां. अं. हिं वि., वर्धा (महाराष्ट्र)

सारांश–
प्रस्तुत शोधपत्र के माध्यम से हिंदी पाठ के संवेदना विश्लेषण (भाव–विश्लेषण) हेतु एक संगणकीय मॉडल प्रस्तावित किया गया है जो हिंदी भाषा के किसी दिये गये पाठ को वाक्य के स्तर पर स्वचालित रूप से संसाधित एवं विश्लेषित कर भाव (सकारात्मक, नकारात्मक या निष्पक्ष) के आधार पर विभाजित करेगा। उक्त संगणकीय मॉडल के अंतर्गत नवीन प्रविधियों का प्रयोग किया गया है जिससे भाव–विश्लेषण प्रणाली की शुद्धता एवं गुणवत्ता में वृद्धि होगी।

मुख्य शब्द (Keywords) रू भाव–विश्लेषण, प्राकृतिक भाषा संसाधन, सोशल मीडिया, हिंदी भाषा, यूनिकोड।

1.विषय परिचय

1.1.भाव–विश्लेषण (Sentiment Analysis)

भाव–विश्लेषण एक संगणकीय प्रक्रिया है जिसका प्रयोग मुख्य रूप से किसी दिये गये पाठ (Text) को उसमें उपस्थित भावयुक्त शब्दों या पदों के आधार पर विश्लेषित एवं संसाधित कर सकारात्मक, नकारात्मक या निष्पक्ष के रूप में वर्गीकृत करने हेतु किया जाता है[3]। इस प्रक्रिया के माध्यम से किसी प्रयोगकर्ता(व्यक्ति, वक्ता या उपभोक्ता) के द्वारा किसी विषय(उत्पाद, सेवाओं, मुद्दों आदि) के प्रति इंटरनेट अनुप्रयोगों (सोशल मीडिया, डिस्कशंस फोरम्स, रिव्यू चैनल्स, आदि) पर व्यक्त प्रतिक्रियाओं का विश्लेषण किया जाता है।

भाव–विश्लेषण में किसी पाठ को निम्न तीन प्रकार से विभाजित किया जाता है–

1)सकारात्मक (Positive) –ऐसी प्रतिकिया जिसमे वक्ता के सकारात्मक भाव उपस्थित हों, सकारात्मक प्रतिक्रिया कहलाती है।
उदाहरण– विश्वकप में भारत ने शानदार प्रदर्शन किया।

2)नकारात्मक (Negative) –ऐसी प्रतिकिया जिसमे वक्ता नकारात्मक भाव उपस्थित हों, नकारात्मक प्रतिक्रिया कहलाती है।
उदाहरण– मोहन ने खराब गेंदबाजी की।

3)निष्पक्ष (Neutral)– ऐसी प्रतिकिया जिसमे केवल सामान्य सूचनाएं हों तथा सकारात्मक या नकारात्मक भाव अनुपस्थित हों, निष्पक्ष प्रतिक्रिया कहलाती है।
उदाहरण– रेल बजट की घोषणा हो चुकी है।

1.2.भाव–विश्लेषण के विभिन्न स्तर (Different levels of Sentiment Analysis)

स्तर	कार्य	उदाहरण
शब्द या पद स्तरीय	इस स्तर के भाव–विश्लेषण में दिये गए पाठ में उपस्थित सभी भावयुक्त शब्दों/पदों को चिन्हित कर प्रत्येक को उसके ध्रुवण वर्ग(सकारात्मक, नकारात्मक या निष्पक्ष) के अनुसार विभाजित किया जाता है ।	खराब(**नकारात्मक**), जीत(**सकारात्मक**)
वाक्य स्तरीय	इस स्तर के भाव–विश्लेषण में दिये गए पाठ में से सर्वप्रथम वाक्यों को पृथक किया जाता है। इसके पश्चात प्रत्येक वाक्य को अलग–अलग विश्लेषित कर उसके ध्रुवण वर्ग(सकारात्मक, नकारात्मक या निष्पक्ष) के अनुसार विभाजित किया जाता है ।	भारत ने खराब प्रदर्शन किया। (**नकारात्मक**) इसके बावजूद उसने मैच जीत लिया(**सकारात्मक**)
दस्तावेज़ स्तरीय	इस स्तर के भाव–विश्लेषण में दिये गए संपूर्ण पाठ को विश्लेषित कर केवल एक ध्रुवण स्तर(सकारात्मक,नकारात्मक या निष्पक्ष) पर विभाजित किया जाता है ।	भारत ने खराब प्रदर्शन किया। इसके बावजूद उसने मैच जीत लिया। (**सकारात्मक**)

1.3.भाव–विश्लेषण के अनुप्रयोग

- व्यावसायिक रणनीति निर्माण में।
- निर्णय सहायक प्रणाली के रूप में।
- उत्पाद विश्लेषण में।
- मशीनी–अनुवाद में।

1.4.हिंदी पाठ का भाव–विश्लेषणरू महत्व एवं उपयोगिता (Significance & Utilization)

विश्व में सर्वाधिक समझी, बोली एवं प्रयोग की जाने वाली भाषाओं में हिंदी भाषा का द्वितीय स्थान है[4] इसलिये सूचना प्रौद्योगिकी के माध्यम से इंटरनेट पर हिंदी के वैश्विक प्रसार एवं प्रचार हेतु विभिन्न प्रकार के मानक एवं टूल्स निर्मित किये जा रहे हैं। हिंदी भाषा के लिए यूनिकोड मानक(UTF–8) परिभाषित होने के पश्चात हिंदी भाषा आधारित इंटरनेट अनुप्रयोगों(सोशल मीडिया, ब्लाग्स, वेबसाईट्स आदि) की संख्या में तीव्र गति से वृद्धि हुई है। माइक्रोसॉफ्ट एवं गूगल आदि कंपनियों ने विभिन्न भाषाओं में टाईपिंग को सर्वसुलभ एवं आसान बनाने हेतु कई प्रकार के ट्रांसलिटेरेशन या फोनेटिक आधारित टाइपिंग टूल्स विकसित किये हैं। इन टूल्स की सहायता से कोई व्यक्ति बड़ी ही सरलता से अपनी विचारों या सूचनाओं को अपनी ही भाषा में कंप्यूटर पर संसाधित या संग्रहित कर सकता है। हिंदी भाषा के लिये भी इसी प्रकार के टूल्स विकसित किये गये हैं जिसके कारण वर्तमान में सोशल मीडिया का उपयोग बढ़ा है और इंटरनेट पर हिंदी सूचनाओं या प्रतिक्रियाओं की उपलबद्धता बढ़ी है[6]। इन सूचनाओं या प्रतिक्रियाओं के उचित विश्लेषण की आवश्यकता है जिससे यह किसी संस्था, व्यक्ति या शोधकर्ता के लिये महत्वपूर्ण व उपयोगी सिद्ध हो सके एवं इसके आधार पर कोई महत्वपूर्ण निर्णय लिया जा सके।

1.5.उद्देश्य (Objectives)

1.हिंदी पाठ भाव–विश्लेषण की शुद्धता एवं गुणवत्ता में वृद्धि करना।
2.भाव–विश्लेषण हेतु नवीन एवं अपेक्षाकृत प्रभावी अभिगम की खोज करना।
3.हिंदी पाठ भाव–विश्लेषण हेतु संगणकीय प्रणाली विकसित करना।

2.संबंधित कार्य (Related Work)

हिंदी भाषा के भाव–विश्लेषण संबंधी शोधकार्य अन्य भाषाओं (विशेषकर अंग्रेजी) की अपेक्षाकृत कम हुए हैं। अभी तक हिंदी पाठ के भाव–विश्लेषण के क्षेत्र में किए गए शोधकार्यों में से कुछ प्रमुख निम्न हैं–

1.अक्षत बक्लीवाल, पीयूष अरोरा (IIIT, हैदराबाद) के द्वारा Hindi subjective Lexicon (HSL) का निर्माण किया गया है, जिसमें भावयुक्त शब्दों से मिलते जुलते सभी संभव (Synonyms) तथा विपरीतार्थक (Antonyms) शब्दों को संकलित कर वर्डनेट का प्रयोग किया गया है।[1]

2.आदित्य जोशी, बालामुरली तथा पुष्पक भट्टाचार्य (IIT] Bombay) के द्वारा हिंदी भाषा के भाव–विश्लेषण हेतु Hindi&SentiWordNet (H&SWN) की प्रविधि प्रस्तुत की गई है जिसमें समस्त भावयुक्त पदों (Sentiment&bearing Phrases) का लेक्सिकल कार्पस बनाया गया है।[2]

3.शोध–प्रविधि (Research Methodology)

3.1.सामग्री संकलन (Data Collection)

प्रस्तुत शोधकार्य हेतु हिंदी के कुछ प्रमुख वेबसाइट्स, ब्लाग्स एवं ई–न्यूज पेपर में से डाटा का संकलन किया जाएगा एवं संकलित डाटा का एक कार्पस बनाया जाएगा।

3.2. संसाधन निर्माण (Resource Genaration)

हिंदी भाषा को शोध–संसाधन की दृष्टि से अल्प–संसाधनीय भाषा (Scarce Resource Language) माना जाता है क्योंकि हिंदी भाषा के भाव–विश्लेषण संबंधित सभी शोध–कार्य अभी विकासशील अवस्था में है अत: हिंदी पाठ के भाव–विश्लेषण प्रक्रिया हेतु सभी आवश्यक संसाधनों का निर्माण या संग्रहण भी एक चुनौती है। प्रस्तुत शोधकार्य के अंतर्गत निम्न संसाधनों का विकास या संग्रहण किया गया है–

3.2.1. भाव–कोश निर्माण (Sentiment Lexicon Generation)

भाव–विश्लेषण प्रणाली की गुणवत्ता एवं शुद्धता में वृद्धि करने हेतु एक भाव–कोश का निर्माण किया जाएगा जिसके अंतर्गत हिंदी भाषा के समस्त भावयुक्त शब्दों (विशेषण, क्रिया–विशेषण आदि) को संग्रहित किया जाएगा। भाव–कोश में प्रत्येक शब्द को एक निश्चित ध्रुवण वर्ग(सकारात्मक, नकारात्मक या निष्पक्ष) में परिभाषित कर एक निश्चित आंकिक मान प्रदान किया जाएगा।

Table 1 : **हिंदी भाव–कोश**

ID	Sentiment Word	Polarity	Score
1.	अच्छा	Positive	.40
2.	बेहतर	Positive	.60
3.	जबरदस्त	Positive	.80
4.	खराब	Negative	–.40
5.	बदतर	Negative	–.60

3.2.2. शब्द–भेद टैग (POS Tag)

किसी भी भाषा के शब्दों की व्याकरणिक कोटि ज्ञात करने हेतु एक टैगसेट की आवश्यकता होती है इसलिये प्रस्तावित संगणकीय प्रणाली के अंतर्गत एक टैगसेट का निर्माण किया जाएगा जो TDIL द्वारा हिंदी भाषा के लिये निर्धारित किए गए मानक टैगसेट (Unified Parts of Speech (POS) Standard in Indian Languages&Draft Standard &Version 1.0) पर आधारित होगा।

3.2.3. नियम विकास (Rule Development)

प्रस्तुत संगणकीय मॉडल हेतु कुछ नियमों का संग्रहण या प्रतिपादन किया गया है जिनमें से कुछ प्रमुख है–

3.2.3.1. शब्द–भेद टैगिंग हेतु

निवेशित पाठ में उपस्थित समस्त शब्दों की व्याकरणिक कोटियों की पहचान करने हेतु नवनीत गर्ग, विशाल गोयल एवं सुमन प्रीत[5] द्वारा प्रतिपादित किये गये नियमों का प्रयोग किया जाएगा।

3.2.3.2. ध्रुवणता नियंत्रण हेतु

पाठ या वाक्य में उपस्थित कुछ नकारात्मक शब्दों(ना, नही) के कारण कभी–कभी वाक्य की ध्रुवणता परिवर्तित(सकारात्मक से नकारात्मक या नकारात्मक से सकारात्मक) हो जाती है इसलिए ध्रुवणता परिवर्तन (Polarity Shifting) की इस स्थिति को नियंत्रित करने हेतु नियमों का विकास गया है जो निम्नलिखित हैं–

A. विशेषण की स्थिति में.

1.किसी वाक्य में यदि विशेषण(सकारात्मक) उपस्थित पश्चात हो तो उस वाक्य की सकारात्मक होने की प्रायिकता (Probability) अधिक होती है।
2.किसी वाक्य में यदि विशेषण(सकारात्मक) के पश्चात नकारात्मक शब्द उपस्थित हो तो उस वाक्य की नकारात्मक होने की प्रायिकता बढ़ जाती है।
3.किसी वाक्य में यदि विशेषण(नकारात्मक) उपस्थित हो तो उस वाक्य की नकारात्मक होने की प्रायिकता बढ़ जाती है।
4.किसी वाक्य में यदि विशेषण(नकारात्मक) के पश्चात नकारात्मक शब्द उपस्थित हो तो उस वाक्य की सकारात्मक होने की प्रायिकता बढ़ जाती है।

B. क्रिया–विशेषण की स्थिति में.

5.किसी वाक्य में यदि क्रिया–विशेषण (सकारात्मक) के पश्चात नकारात्मक शब्द उपस्थित हो तो उस वाक्य की नकारात्मक होने की प्रायिकता सबसे अधिक होती है।
6.किसी वाक्य में यदि क्रिया–विशेषण (सकारात्मक) उपस्थित हो तो उस वाक्य की सकारात्मक होने की प्रायिकता सबसे अधिक होती है।
7.किसी वाक्य में यदि क्रिया–विशेषण (नकारात्मक) उपस्थित हो तो उस वाक्य की नकारात्मक होने की प्रायिकता अधिक होती है।
8.किसी वाक्य में यदि क्रिया–विशेषण (नकारात्मक) के पश्चात नकारात्मक शब्द उपस्थित हो तो उस वाक्य की सकारात्मक होने की प्रायिकता बढ़ जाती है।

4. भाव–विश्लेषक (Sentiment Analyzer)

हिंदी पाठ के भाव–विश्लेषण हेतु प्रस्तावित संगणकीय प्रणाली को 'भाव–विश्लेषक' का नाम दिया है। यह प्रणाली यूनिकोड आधारित देवनागरी लिपि पर ही क्रियान्वित होगा। विंडो आधारित इस प्रणाली को हिंदी भाषा के लिये विकसित जा रहा है लेकिन प्रणाली संचालन हेतु उपयोगकर्ता को द्विभाषी इंटरफेस(हिंदी एवं अंग्रेजी) की सुविधा प्रदान की जाएगी। इस प्रणाली के द्वारा निवेशित पाठ को शब्द एवं वाक्य के स्तर पर विश्लेषित किया जाएगा।

4.1. कार्यविधि (Working Procedure)

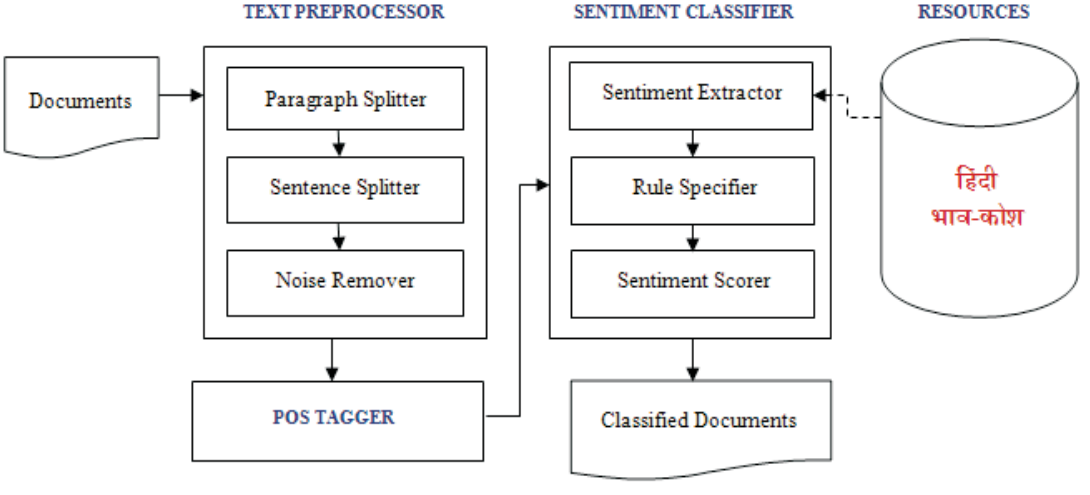

Figure 1: भाव–विश्लेषक (कार्यविधि)

प्रस्तावित संगणकीय प्रणाली निम्न चरणों में क्रियान्वित होगी :

चरण 1: पाठ निवेशन (Data Input)
प्रथम चरण में भाव–विश्लेषण किये जाने वाले पाठ का प्रणाली में निवेश किया जाएगा।

चरण 2 : डाटा पूर्व–संसाधन (Data Preprocessing)
निवेशित पाठ को वाक्य के स्तर पर पृथक किया जाएगा क्योंकि प्रणाली द्वारा पाठ का वाक्य स्तरीय भाव–विश्लेषण किया जाएगा। तत्पश्चात प्रत्येक वाक्य में से अवांछनीय तत्व(जैसे– विशेष अक्षर, चिन्ह, लोगो, इमोटिकोंस आदि) को विलोपित किया जाएगा।

चरण 3: शब्द–भेद टैगिंग (POS Tagging)
पाठ में से प्रत्येक शब्द–भेद को उसके भेद के अनुसार चिन्हित किया जाएगा।

चरण 4: भाव–विश्लेषण (Sentiment Analysis)
यह प्रणाली का सबसे महत्वपूर्ण चरण है। इस चरण में भावयुक्त शब्दों का भाव–कोश के अनुसार आंकिक मान ज्ञात किया जाएगा तत्पश्चात पाठ का विश्लेषण किया जाएगा। अंत में सभी अंको का योग किया जाएगा। यदि आंकिक योग सकारात्मक हो तो पाठ की ध्रुवणता सकारात्मक एवं आंकिक योग नकारात्मक हो तो पाठ की ध्रुवणता नकारात्मक निश्चित की जाएगी।

चरण 5: परिणाम (Result)
अंतिम चरण में प्रणाली द्वारा परिणाम प्रदर्शित किया जाएगाद्य

5.निष्कर्ष एवं भविष्य की संभावनाएं

प्रस्तावित संगणकीय प्रणाली हिंदी भाषा के भाव–विश्लेषण हेतु उपयोगी सिद्ध होगी। प्रस्तावित प्रणाली केवल यूनिकोड आधारित देवनागरी लिपि पर ही कार्य करेगी। हिंदी भाषा की रुपिमिक विशेषता, मुक्त शब्द क्रम एवं विभिन्न वाक्य संरचना के हिंदी पाठ का भाव–विश्लेषण करना सरल नहीं है। इस शोधकार्य के अंतर्गत भावयुक्त शब्दों की श्रेणी में केवल विशेषण, क्रिया–विशेषण को सम्मिलित किया गया है किंतु प्रयोगकर्ता के द्वारा कभी–कभी बहुशब्दीय अभियक्तियों(मुहावरों, कहावतों एवं लोकोक्तियों) का उपयोग किसी विषय के प्रति अपने सकारात्मक/ नकारात्मक भाव को व्यक्त करने के लिये किया जाता है इसलिए नियम–समूह एवं भाव–कोश में विस्तार से अपेक्षाकृत बेहतर परिणाम प्राप्त हो सकेंगे एवं प्रणाली की गुणवत्ता एवं शुद्धता में वृद्धि होगी।

6.संदर्भ–सूची (References)

[1]Arora] Piyush- (2013)- Sentiment Analysis for Hindi Language (MS Thesis)- IIIT Hyderabad-
[2]Joshi] A-] A- R- Balamuraly and Pushpak Bhattacharya. (2010).A Fall&back Strategy for Sentiment Analysis in Hindi: a Case Study- Proceedings of ICON: 8th International Conference on Natural Language Processing] Macmillan Publishers] India-
[3]Pang B-] and Lillian Lee- (2008)- Opinion Mining and Sentiment Analysis- Foundations and Trends in Information Retrieval 2(1&2): 1&135.
[4]http://en-wikipedia-org/wiki/List of languages by number of native speakers#More than 100 million native speakers.
[5]Garg] Navneet] Vishal Goyal and Suman Preet- (2012). Rule Based Hindi Part of Speech Tagger- Proceedings of COLING 2012: Demonstration Papers] pages 163&174-
[6]पांचजन्य, जुलाई 13, 2014. कम्प्यूटर पर आईटी ने खोले भारतीय भाषाओं के रास्ते. पृष्ठ संख्या 30–31.

Publish Research Article
International Level Multidisciplinary Research Journal
For All Subjects

Dear Sir/Mam,

We invite unpublished Research Paper,Summary of Research Project,Theses,Books and Books Review for publication,you will be pleased to know that our journals are

Associated and Indexed,India

- Directory Of Research Journal Indexing
- International Scientific Journal Consortium Scientific
- OPEN J-GATE

Associated and Indexed,USA

- DOAJ
- EBSCO
- Crossref DOI
- Index Copernicus
- Publication Index
- Academic Journal Database
- Contemporary Research Index
- Academic Paper Databse
- Digital Journals Database
- Current Index to Scholarly Journals
- Elite Scientific Journal Archive
- Directory Of Academic Resources
- Scholar Journal Index
- Recent Science Index
- Scientific Resources Database

Review Of Research Journal
258/34 Raviwar Peth Solapur-413005,Maharashtra
Contact-9595359435
E-Mail-ayisrj@yahoo.in/ayisrj2011@gmail.com
Website : www.ror.isrj.org