Chapter 5
Conversion of Higher into Lower Language Using Machine Translation

Raghvendra Kumar
Lakshmi Narain College of Technology, India

Prasant Kumar Pattnaik
KIIT University, India

Priyanka Pandey
Lakshmi Narain College of Technology, India

ABSTRACT

This chapter addresses an exclusive approach to expand a machine translation system beginning higher language to lower language. Since we all know that population of India is 1.27 billion moreover there are more than 30 language and 2000 dialects used for communication of Indian people. India has 18 official recognized languages similar to Assamese, Bengali, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Tamil, Telugu, and Urdu. Hindi is taken as regional language and is used for all types of official work in central government offices. Commencing such a vast number of people 80% of people know Hindi. Though Hindi is also regional language of Jabalpur, MP, India, still a lot of people of Jabalpur are unable to speak in Hindi. So for production those people unswerving to know Hindi language we expand a machine translation system. For growth of such a machine translation system, used apertium platform as it is free/open source. Using apertium platform a lot of language pairs more specifically Indian language pairs have already been developed. In this chapter, develop a machine translation system for strongly related language pair i.e Hindi to Jabalpuriya language (Jabalpur, MP, India).

INTRODUCTION

Natural Language Processing

Natural language is an integral part of our day today lives. Language is the most common and most ancient way to exchange information among human beings. people communicate and record information.NLP is a field of computer science and artificial intelligence; here natural language means the language used by human being for communication among themselves. NLP is a form of human-to-computer interaction. ie the nlp basically implies making human to machine interaction easy and in human language .

One of the major problems encountered by any nlp system is lexical ambiguity here the term lexical ambiguity means the particular word having more than one meaning lexical ambiguity in simple words can better be stated as presence of homonymy and polysemy. Ambiguity is further of two types syntactic and semantic. Syntactic ambiguity means when sentence can be parsed in more than one manner. The word's syntactic ambiguity can be resolved in language processing by part-of-speech taggers with very high level of accuracy. Semantic ambguty The problem of resolving semantic ambiguity is generally known as word sense disambiguation (WSD) and has been proved to be more difficult than syntactic disambiguation.

Word Sense Disambiguation

Any natural language known to human beings there exists that can have more than one possible meaning, for example a bat can be a small creature or a cricket equipment and a bank can mean river 's bank or a money bank (the financial institution one) . since if the other meaning than the intended one is used in particular context it can cause a huge translation hazard Given these complications, it is important for a computer to correctly determine the meaning in which a word is used. hence it is very clear to s that Ambiguity is natural phenomena in to human language and it constitutes one of the most important problem for computational applications in of Natural Language Processing (NLP). Ideally, systems should be able to deal with ambiguity in order to increase performance in nlp applications such as Text Summarization and Information Retrieval. The process of assigning the correct sense of ambiguous words Words can have different senses. Some words have multiple meanings. This is called Polysemy. For example: bank can be a financial institute or a river shore. Sometimes two completely different word are spelled the same.

For example: Can, can be used as model verb: You can do it, or as container: She brought a can of soda. This is called Homonymy. Distinction between polysemy and homonymy is not always clear. Word sense disambiguation (WSD) is the

Related Content

Hybrid Query and Data Ordering for Fast and Progressive Range-Aggregate Query Answering

Cyrus Shahabi, Mehrdad Jahangiriand Dimitri Sacharidis (2005). *International Journal of Data Warehousing and Mining (pp. 49-69).*

www.irma-international.org/article/hybrid-query-data-ordering-fast/1751

Graph Mining and Its Applications in Studying Community-Based Graph under the Preview of Social Network

Bapuji Raoand Anirban Mitra (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 970-1022).*

www.irma-international.org/chapter/graph-mining-and-its-applications-in-studying-community-based-graph-under-the-preview-of-social-network/150202

Social Media Mining: A New Framework and Literature Review

Vipul Guptaand Mayank Gupta (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 2401-2414).*

www.irma-international.org/chapter/social-media-mining/150271

Finding Non-Coincidental Sporadic Rules Using Apriori-Inverse

Yun Sing Koh, Nathan Rountreeand Richard O'Keefe (2006). *International Journal of Data Warehousing and Mining (pp. 38-54).*

www.irma-international.org/article/finding-non-coincidental-sporadic-rules/1765

Traffic Safety Implications of Travel Demand Management Policies: The Cases of Teleworking and Fuel Cost Increase

Ali Pirdavani, Tom Bellemans, Tom Brijs, Bruno Kochanand Geert Wets (2014). *Data Science and Simulation in Transportation Research (pp. 115-140).*

www.irma-international.org/chapter/traffic-safety-implications-of-travel-demand-management-policies/90069