सारांश

प्रस्तुत शोध यौगिक संज्ञा पर निर्धारित है। यौगिक संज्ञा दो या दो से अधिक शब्दों का योग होता है, जिसमें एक संज्ञा का योग होना आवश्यक होता है तथा इसमें शब्दों के सिर्फ मूल रूप का इस्तेमाल होता है। इस शोध के दौरान एक टूल बनाया गया है जो एक संज्ञा को इनपुट करने पर उस संज्ञा से बनने वाले सभी यौगिक संज्ञाओं तथा उससे जुड़ी सूचनाएं प्रदर्शित करता है। इस टूल को बनाने के लिए संकर तकनीक का प्रयोग किया गया। इस टूल के लिए नियमों का विकास किया गया है। ये नियम शब्दों के विश्लेषण के पश्चात बनाये गए हैं। इसमें यह देखा गया कि यौगिक संज्ञा बनाने वाले शब्दों के बीच कौन से गुण एक जैसे हैं और कौन से गुण अलग-अलग हैं, इन गुणों के आधार पर नियम बनायें गये हैं।

जैसा की इस शोध के दौरान विशेषज्ञ से मिली टिप्पणियों के आधार पर नियमों कि सहायता से टूल को चलाने कि कोशिश की गयी जो असफल रही। इसमें देखा गया कि नियमों के आधार पर यह सिर्फ कुछ शब्दों के ऊपर सही परिणाम सक्षम है, यहाँ तक की समान अर्थ वाले अन्य शब्द के आने पर भी यह सही परिणाम उत्पादित नहीं कर पाता है। इस समस्या को दूर करने के लिए एक डाटाबेस का निर्माण किया गया, यह डाटाबेस नियमों की सहायता से यौगिक संज्ञा का निर्माण करता है। इस शोध के दौरान एक डाटाबेस का निर्माण किया गया जिसके लिए एक टैगसेट का निर्माण किया गया। यह टैगसेट डाटाबेस में शब्दों के ontological features को संग्रहीत करने में इस्तेमाल की जाती है।

इस शोध के दौरान यह पता चला की एक समान अर्थ रखने वाले शब्द भी एक ही शब्द के साथ यौगिक बनाने में सक्षम नहीं हैं। ये परिस्थिति ज्यादातर तब आती है जब कोई शब्द किसी अन्य भाषा का हो और वह हिंदी में प्रयोग होने लगा हो, जैसे – अरबी, अंग्रेजी आदि भाषाओं के शब्द।

इसके अलावा यह भी देखा गया कि समान प्रकृति के शब्द जो कि एक से भाषा के हो और उनका ontological features भी समान प्रकार का हो, ऐसी परिस्थिति में भी ज्यादातर जगहों पर समान नियम सही परिणाम नहीं देते हैं।

संज्ञाओं के अन्य भाषिक व्याकरणिक इकाई की अपेक्षा काफी अधिक संख्या में होने के कारण इसके डाटाबेस निर्माण में मुश्किलों का सामना करना पड़ा। हिंदी भाषा में अन्य भाषाओं के काफी शब्दों का प्रयोग होने के कारण यह समस्या और जटिल हो गयी।

नियम निर्माण के समय प्रत्येक दूसरे-तीसरे शब्दों के ऊपर नियम प्रयोग के बाद सही परिणाम ना देने के कारण कई नियमों का निर्माण करना पड़ा। कई जगहों पर सिर्फ एक शब्द पर अलग नियम का प्रयोग करना पड़ा।

यह टूल हिंदी आधारित है अतः यह हिंदी के विकास में सहायक होगा।

Summary

This research is based on Compound Noun. It has been developed for Hindi Language. In Compound Noun it is the sum of two or more words, in which it is necessary to have a noun and it's made of root word. During this research a tool has been developed. It takes an input in the form of noun and it give us compound noun and the information of both words. This tool is based on hybrid approach in which a database and rules are developed. The rules is based on the analysis of word which is known as ontological features. After analysing the word, the features is the base of rule.

At start we try to develop the tool on the base of rule but that was not successful. It is known that only few word can be generated by rules. Even after same meaning word cannot produce right results. To solve this problem a database has been developed. In this database we developed a tag set on which the database is created. The database is made for storing ontological features and information of word.

During this research it's known that if a word is from different language even they have same meaning they are not eligible for making compound noun. The word from other languages word which used in Hindi language are – English, Arabic, etc.

Even the word of same nature which have same ontological features cannot be guaranteed that they produce same compound noun, in different situations most of the times they don't give right result.

During rules development process a new rule has been created for every second or third word.

This tool is developed for Hindi and it will help in development of Hindi.