मशीनी अनुवाद
Machine Translation

यूजीसी सीबीसीएस स्कीम के तहत बीए (संस्कृत) के एईईसी-3 (AEEC-3) के पाठ्यक्रम पर आधारित

सुभाष चन्द्र

विद्यानिधि प्रकाशन
दिल्ली

विषयसूची (TABLE OF CONTENTS)

विषय-सूची ..... i-ii
Table of Contents

पुस्तक के विषय में ..... iii-v
About the Book

आभार ..... vi-vii
Preface

प्रथम अध्याय 001-084
मशीनी अनुवाद का सामान्य परिचय एवं इतिहास
Chapter One
Brief Introduction and History of Machine Translation

द्वितीय अध्याय 085-119
मशीनी अनुवाद का सर्वेक्षण
Chapter Two
Survey of Machine Translation

तृतीय अध्याय 120-148
मशीनी अनुवाद की विभिन्न विधियां
Chapter Three
Machine Translation (MT) Approaches

चतुर्थ अध्याय 149-241
मशीनी अनुवाद में चुनौतियां
Chapter Four
Challenges in Machine Translation

सन्दर्भग्रन्थसूची 319-320

References

प्रथम अनुक्रमणिका.....321-344
भारतीय भाषाओं के लिये विकसित एमटी सिस्टम की सूची
Appendix One
List of Developed MT Systems for Indian Languages

द्वितीय अनुक्रमणिका.....321-344
भारतीयेतर भाषाओं के लिये विकसित एमटी सिस्टम की सूची
Appendix Two
List of Developed MT Systems for Other than Indian Languages

प्रथम अनुक्रमणिका.....321-344
मशीनी अनुवाद पर कार्य करने वाले संस्थानों की सूची
Appendix Three
List of the Institute working on Machine Translation

लेखक के विषय में.....350-352
About the Author

पुस्तक के विषय में

भाषाविज्ञान संसार की किसी भी भाषाओं का वैज्ञानिक अध्ययन करता है। सूचना एवं प्रसारण के इस युग में सभी प्रकार की विधाओं का डिजटलीकरण हो रहा है। सूचनाओं का आदान-प्रदान भाषा के ही माध्यम से होता है। सूचनाओं को सभी तक तत्काल पहुंचाने के लिये कम्प्यूटर का प्रयोग बहुत ही तेजी से हो रहा है। अतः भाषाविज्ञान के क्षेत्र में कम्प्यूटर एवं भाषाविज्ञान को एक साथ जोड़कर एक नये अन्तर्विषयक अनुसंधान का उदय 1950 के दशक में हुआ। जिसको संगणकीय भाषाविज्ञान के नाम से जाना गया। जिसका उद्देश्य एक भाषा का दूसरे भाषा में ऑटोमैटिक मशीनी अनुवाद था। धीरे-धीरे इसका विकास हुआ और बहुत सारे शोधकर्त्ताओं ने इस विषय पर कार्य करना शुरु किया। भारत में इसकी शुरुआत 1980 के दशक में हुई। संस्कृत भाषा का कम्प्यूटर के लिये महत्त्व की चर्चा 1985 में नासा के वैज्ञानिक रिक ब्रिग्स के एक शोध से सामने आई। धीरे-धीरे संगणकीय भाषाविज्ञान का विकास भारत में भी हुआ और सूचना प्रौद्यिगिकी मन्त्रालय, भारत सरकार ने भी 1991 में भारतीय भाषाओं के लिये तकनीक के विकास के लिये एक विभाग भी बनाया। यह विभाग भारतीय भाषाओं के लिये तकनीक विकास के लिये अनुदान आदि प्रदान करके तकनीक का विकास कार्य करता है। तथा तकनीक को लोगों तक पहुंचाने का कार्य भी करता है। संगणकीय भाषाविज्ञान आज लगभग सभी आईआईटी, आईआईआईटी, विश्वविद्यालयों तथा अन्य अनुसंधान केन्द्रों में पढ़ाया या शोध कराया जाता है। निजी कम्पनियां जैसे गूगल, माइक्रोसॉफ्ट, याहू, सैमसंग एवं अन्य बहुत सारी ने भी इस क्षेत्र में कार्य प्रारम्भ कर दिया है।

मशीनी अनुवाद भी संगणकीय भाषाविज्ञान का एक उपविषय है। जिसमें मशीनी अनुवाद के किन-किन तकनीकों का प्रयोग किया जाता है आदि विभिन्न विषयों पर शोध कार्य किये जाते हैं। मशीनी अनुवाद नामक यह पुस्तक उन सभी छात्रों, जिज्ञासुओं तथा शोध कर्त्ताओं के लिये लिखी गयी है जो मशीनी अनुवाद के क्षेत्र में हिन्दी माध्यम से प्रवेश पाना चाहते हैं। यह पुस्तक भारत के सभी केन्द्रीय विश्वविद्यालयों में स्नातक स्तर पर यूजीसी द्वारा 2015 में प्रारम्भ किये गए सीबीसीएस स्कीम के तहत बीए (संस्कृत) के एईईसी (AEEC)-3 "मशीनी अनुवाद: उपकरण एवं तकनीक" नामक पाठ्यक्रम के सभी विषयों को समाहित करती है। इसकी भाषा बहुत ही आसान एवं ग्राह्य है साथ ही साथ अंग्रेजी भाषा के किसी भी शब्दावली का हिन्दी अनुवाद नही किया गया है उन्हें

केवल देवनागरी में लिप्यन्तरित किया गया है साथ ही साथ कोष्ठक में अंग्रेजी भी लिखी गई है।

अकेले दिल्ली विश्वविद्यालय से सम्बद्ध लगभग 45 महाविद्यालयों में संस्कृत पढाई जाती है। इसके साथ ही साथ लगभग सभी भारतीय विश्वविद्यालयों में संस्कृत पढाई जाती है। दिल्ली विश्वविद्यालय में स्नातक प्रथम वर्ष में हर वर्ष लगभग 2000-2500 विद्यार्थी प्रवेश लेते हैं। सीबीसीएस स्कीम के तहत इस पाठ्यक्रम को किसी भी विषय का छात्र चयन कर सकता है चाहे वह संस्कृत का हो या न। यह विषय नया है इस पर कोई भी पुस्तक हिन्दी भाषा में उपलब्ध नहीं है। अतः यह पुस्तक इन सभी विद्यार्थियों के लिये बहुत ही उपयोगी सिद्ध होगी । इस पुस्तक का अध्ययन करने के बाद विद्यार्थी/शोधार्थी मशीनी अनुवाद की सामान्य अवधारणा से परिचित तो होगा ही साथ ही साथ इसके इतिहास एवं इसमें शोध करने के लिये प्रयुक्त विधियों के विषय में भी अवगत होगा। इस पुस्तक में मशीनी अनुवाद से सम्बन्धित कार्य करने वाले प्रमुख अनुसंधान केन्द्रों तथा प्रमुख मशीनी अनुवाद सिस्टमों का भी परिचय गया है। उत्तरी भारत में स्थित लगभग सभी शैक्षिक संस्थानों में संस्कृत विषय हिन्दी माध्यम में पढ़ाया जाता है। अतः यह पुस्तक हिन्दी भाषा में लिखी गई है जिससे हिन्दी माध्यम के विद्यार्थी इस विषय में प्रवेश आसानी से पा सकें। इस क्षेत्र के सभी तकनीकी शब्द अंग्रेजी में समझने आसान होते हैं इसलिये ऐसे किन्ही भी शब्दों का हिन्दी अनुवाद नहीं किया गया है बल्कि उसे देवनागरी में लिख दिया गया है और कोष्ठक में उसको रोमन में भी लिखा गया है।

इस पुस्तक को लेखन की प्रेरणा उस समय मिली जब 2014 में यूजीसी के लिये सीबीसीएस (Choice based Credit System) पाठ्यक्रम बनाया जा रहा था। सीबीसीएस के तहत संस्कृत के बीए (ऑनर्स) के एईईसी-3 मशीनी अनुवादः टूल्स एवं तकनीक (Machine Translation: Tools and Technique) नामक एक पाठ्यक्रम शामिल किया गया। सीबीसीएस स्कीम के तहत आने वाले लगभग सभी पाठ्यक्रम नये हैं तथा यूजीसी द्वारा सभी केन्द्रीय विश्वविद्यालयों एवं इनसे सम्बद्ध महाविद्यालयों में 2015 से लागू कर दिए गए हैं। इनमें से बहुत से प्रश्नपत्रों के लिये आज तक कोई भी पाठ्य-पुस्तक उपलब्ध नही है। दिल्ली विश्वविद्यालय में यह पाठ्यक्रम 2015 में लागू किया गया। दिल्ली विश्वविद्यालय से सम्बद्ध महाविद्यालयों के छात्रों एवं शिक्षकों से बातचीत के दौरान यह बात सामने आयी कि कुछ प्रश्नपत्रों के लिये हिन्दी माध्यम में पाठ्य-पुस्तकों की महती आवश्यकता है जिनमें यह पाठ्यक्रम भी शामिल है। अतः हमने इस पुस्तक पर कार्य करना प्रारम्भ कर दिया। इस पुस्तक पर कार्य का आरम्भ तो 2015 में ही हो गया था परन्तु इसका प्रकाशन 2017 मार्च के अन्त में हो पाया। सर्वेक्षण के दौरान यह भी पता चला कि अकेले दिल्ली विश्वविद्यालय से सम्बद्ध लगभग 45

महाविद्यालयों में संस्कृत पढाई जाती है। अतः छात्रों की संख्या और भी बढ़ जाती है। जिससे इस पुस्तक की उपयोगिता भी बहुत अधिक बढ़ जाती है। अतः इस पुस्तक पर कार्य प्रारम्भ किया गया। इस पुस्तक के लेखक का इस क्षेत्र में 13 वर्षों का अनुसंधान एवं विकास का अनुभव तथा 5 वर्षों का शिक्षण के अनुभव होने के साथ-साथ दिल्ली विश्वविद्यालय द्वारा सीबीसीएस (संस्कृत) पाठ्यक्रम निर्माण में महती भूमिका अदा करने के कारण यह कार्य अधिक आसान हो गया। इस पुस्तक के लेखन से पूर्व इस विषय पर छात्रों के समक्ष कई बार लेक्चर दिये गये उनकी समस्याओं एवं सुझावों का भरपूर प्रयोग कर यह पुस्तक छात्रों को समझने योग्य बनाने का प्रयास किया गया है। इसकी भाषा बहुत ही आसान एवं ग्राह्य है जो छात्रों एवं शिक्षकों के लिये बहुत ही उपयोगी सिद्ध होगी। क्लासरूम शिक्षण के अनुभवों का भरपूर उपयोग कर यह पुस्तक को एक शिक्षक की शैली में लिखी गई है।

यह पुस्तक कुल चार अध्यायों में विभक्त है। मशीनी अनुवाद का सामान्य परिचय एवं इतिहास नामक प्रथम अध्याय दो खण्डों में विभक्त है। इस अध्याय के प्रथम खण्ड में मशीनी अनुवाद का सामान्य परिचय दिया गया है। जिससे पाठक मशीनी अनुवाद के विषय में जान सकेगा। इसके दूसरे खण्ड में मशीनी अनुवाद का संक्षिप्त इतिहास दिया गया है। इससे छात्र इसके इतिहास से भी अवगत हो सकेगें। पाठ्यक्रम की प्रथम अन्विति इस अध्याय में समाहित है। इसका दूसरा अध्याय मशीनी अनुवाद का सर्वेक्षण है। जो पाठ्यक्रम के दूसरी अन्विति को समाहित करता है। इस अध्याय में मशीनी अनुवाद पर हो रहे समस्त अनुसंधानकार्यों का बड़े ही विस्तार पूर्वक वर्णन करता है। इससे पाठक को भारत ही नही बल्कि अन्य देशों में भी अनुवाद से सम्बन्धित होने वाले कार्यों का सामान्य ज्ञान होगा। यह अध्याय शोधकर्त्ताओं के लिये भी बहुत उपयोगी सिद्ध हो सकता है। तीसरा अध्याय मशीनी अनुवाद में प्रयुक्त होने वाली विभिन्न विधियों का सविस्तार एवं सोदाहरण वर्णन करता है। यह अध्याय पाठ्यक्रम के वर्ग सी के प्रथम यूनिट को समाहित करता है। इसमें नियम आधारित, सांख्यकीय विधि एवं संकर विधि का विशेष रूप से वर्णन किया गया है। सामान्य अवधारणाओं के साथ मशीनी अनुवाद में आने वाली समस्याओं एवं चुनौतियों का वर्णन इस पुस्तक का चौथा एवं अन्तिम अध्याय करता है। यह पुस्तक एक पाठ्यपुस्तक के रूप में प्रयोग के उद्देश्य से लिखी गयी है। फिरभी किसी भी सामान्य जिज्ञासुओं, शोधकर्त्ताओं के लिये भी उपयोगी है। इस पुस्तक में प्रारूप स्वरूप कुछ मशीनी अनुवाद सिस्टमों का विस्तृत परिचय दिया गया है जिनमें से गूगल अनुवाद एवं माइक्रोसॉफ्ट अनुवादक मुख्य हैं। पुस्तक के अन्त में अनुक्रमणी में मशीनी अनुवाद सिस्टम की सूची दी गई है। सब मिलाकर यह पुस्तक देखने के बहुत छोटी है परन्तु मशीनी अनुवाद से सम्बन्धित

सूचनाओं से भरपूर आपके सामने इस उम्मीद से प्रस्तुत है कि आप सभी इससे अवश्य लाभान्वित होंगे और मशीनी अनुवाद की सामान्य अवधारणा से परिचित भी होंगे।

मशीनी अनुवाद के बहुत से छोटे-छोटे घटक होतें हैं । जिन्हें आप दिल्ली विश्वविद्यालय (http://cl.sanskrit.du.ac.in) एवं जवाहरलाल नेहरू विश्वविद्यालय (http://sanskrit.jnu.ac.in) के संस्कृत विभाग की वेबसाइट पर आसानी से देख सकतें हैं। आपके महत्त्वूर्ण सुझाव शायद इस पुस्तक को और भी अधिक अच्छा बना सके। अपने सुझाव ई-मेल के माध्यम से लेखक को **subhash.jnu@gmail.com** पर भेजें।