

भारतीय भाषाओं हेतु डिजिटल संसाधन

धनंजय विलास झाल्टे

zaltedhananjays@gmail.com

भाषा-प्रौद्योगिकी में पी-एच. डी., संप्रति कंसल्टेंट, सी-डैक, पुणे

इंटरनेट के आविष्कार के बाद 'सूचना' और 'संपर्क' के क्षेत्र अद्भुत प्रगति हुई। पारंपरिक सूचना साधनों की जगह इलेक्ट्रॉनिक साधनों ने ली। इलेक्ट्रॉनिक तथा डिजिटल उपकरणों की सहायता से सूचनाओं का आदान प्रदान होने लगा और इसके कारण विश्व के हर क्षेत्र के साथ जुड़ने में मदद हो गई। 'सूचना प्रौद्योगिकी' के कारण 'सूचना' या 'ज्ञान' को विश्व में प्रसारित करने या ज्ञान को अर्जित करने के लिए इनकी बहुत मदद होने लगी और आज यह जरूरत बन गई। यह एक ऐसा ज्ञान का भंडार है, जिसका हम जब चाहे उपयोग कर सकते हैं | इसके माध्यम से कम समय में किसी भी सूचना को पुरे विश्व में फैला सकते है और इसके कारण समय की बचत के साथ-साथ पूरे विश्व, देश, समाज, व्यक्ति के साथ जुड़े रहना आसान हो गया। सूचना-प्रौद्योगिकी ने "ग्लोबल विलेज" की संकल्पना को भी जन्म दिया और भाषा-प्रौद्योगिकी ने 'ई-भाषाई समाज' को जन्म दिया, क्योंकि भाषाई विकसित तकनीकों के कारण भाषाई दूरियाँ खत्म हो गई हैं। आर्थिक उदारतावाद के इस दौर के वैश्विक ग्राम (ग्लोबल विलेज) की संकल्पना संचार प्रौद्योगिकी के कारण सफल हुई है। इस नये युग में ई-कॉमर्स, ई-मेडिसिन, ई-एज्युकेशन, ई-गवर्नंस, ई-बैंकिंग, ई-शॉपिंग, ई-लायब्रेरी, ई-कॉन्फरन्स आदि इलेक्ट्रॉनिक संकल्पनाएँ आ रही हैं। सूचना प्रौद्योगिकी आज शक्ति एवं विकास का प्रतिक बन गई है। कंप्यूटर युग के संचार साधनों में सूचना प्रौद्योगिकी के आगमन से हम लोगों ने 'ई-सूचना समाज' में प्रवेश किया है |

कंप्यूटर और इंटरनेट के आविष्कार के बाद 'भाषाविज्ञान' के क्षेत्र में भी 'बहुत बड़ा बदलाव आ गया। 'भाषा' को 'कंप्यूटर' के साथ जोड़ने से 'भाषा-प्रौद्योगिकी' इस संकल्पना का उदय हुआ। द्वितीय विश्वयुद्ध के

बाद यह महसूस किया गया कि कंप्यूटर को तार्किक नियम दिए जाए तो भाषा-विश्लेषण के कार्य भी किए जा सकते हैं इसलिए इस क्षेत्र में कार्य किए जाने लगे। किंतु शुरू से ही कंप्यूटर और प्रौद्योगिकी की भाषा अंग्रेजी रह चुकी है इसलिए भारत जैसे बहुभाषी देश में अनेक भाषाएँ होने के कारण इन सभी भाषाओं पर कार्य करना और उनको प्रौद्योगिकी के साथ जोड़ना आवश्यकता महसूस होने लगी। भाषिक संप्रेषण के लिए समाज, देश हो या विश्व के लिए बहुभाषिकता यह शुरू से ही एक प्रमुख समस्या रह चुकी है। मनुष्य जन्मतः ही जिज्ञासु होने के कारण ज्ञान के प्राप्ति के लिए हर समय प्रयास रत रहता है, परंतु बहुभाषिक समाज में भाषिक भिन्नता के कारण हर भाषी समाज में स्थित ज्ञान को अर्जित करने में समस्याएँ उत्पन्न होती है। उस ज्ञान को अर्जित करना उस भाषाई समाज के भाषाई ज्ञान के बिना संभव नहीं है। इसलिए इसके परिप्रेक्ष्य में कार्य किए जाने लगे। जिसमें प्राकृतिक भाषा संसाधन के परिप्रेक्ष्य में कार्य किए जाने लगे जिसमें मशीनी अनुवाद, व्याकरण जाँचना, कोश, वर्तनी जाँचना, वाक्-संसाधन (speech processing), ओसीआर, ओएमआर, लिपि परिवर्तन, पद विच्छेदन आदि कार्य कंप्यूटर के माध्यम से किए जाने लगे जिसके कारण आनेवाले समय में बहुभाषिकता और भाषाई दूरियाँ खत्म होने में मदद मिल गई।

भाषा प्रौद्योगिकी के क्षेत्र में मशीनी अनुवाद के इतिहास को देखा जाए तो इस क्षेत्र में द्वितीय विश्वयुद्ध के बाद कार्य का आरंभ किया गया। 1954 में आई.बी.एम. (International Business Machine : IBM), न्यूयार्क आई.बी.एम. द्वारा सिस्ट्रॉन (SYSTRAN) नामक मशीनी अनुवाद प्रणाली का प्रदर्शन किया गया, जो रसायनशास्त्र (Chemistry) के 49 रूसी वाक्यों को अंग्रेजी में अनुवाद करता था जिसकी क्षमता 250 शब्दों और 6 व्याकरणिक नियमों का प्रयोग करना था। शब्दानुवाद के कारण इसमें कई अशुद्धियाँ थी, इसलिए इसे 'Toy' सिस्टम कहा जाने लगा। उस समय जितने भी कार्य किए गए उसके परिणाम आशावर्धक नहीं थे, क्योंकि यह कार्य भाषा के नियमों को ध्यान में रखते हुए नहीं किया गया था जिसके कारण भाषाई अशुद्धियाँ होना स्वाभाविक था। इसके बाद यह बात ध्यान में आ गई की भाषा-अध्ययन के बिना यह कार्य असंभव है। कंप्यूटर की तार्किक प्रणाली का उपयोग भाषिक विश्लेषण, संश्लेषण और संसाधन के लिए किया जा सकता है। इसके बाद 1975-78 में अमेरिका में TINLAP,

TINLAP-2 नामक कार्यशालाएँ हुई जिसका उद्देश्य यह था कि संगणकविज्ञान, भाषाविज्ञान, मनोविज्ञान, गणित, संगणक भाषाविज्ञान, कृत्रिम बुद्धि आदि क्षेत्र में काम करनेवाले वैज्ञानिकों को एक प्लेटफार्म में लाकर उनके लिए एक मंच प्रदान करना था, जिससे इस कार्य में आने वाली समस्याओं को हल किया जाए जिससे यह कार्य करने में सरलता आ जाए। जिसके बाद सारे विश्व और भारत में भी कई महत्वपूर्ण कार्य किए गए।

भारत में प्राकृतिक भाषा संसाधन के क्षेत्र में कार्य करने वाले भारतीय प्रौद्योगिकी संस्थान और वहाँ किए गए कार्य इस प्रकार हैं –

a) IIT कानपुर – ‘आंग्लभारती’ भाषा प्रौद्योगिकी संस्थान, कानपुर १९९१ में प्रो. आर. एम. के० सिन्हा द्वारा आंग्लभारती परियोजना आरंभ की गई। भारत में यंत्रानुवाद क्षेत्र में किया जाने वाला यह पहला कार्य था। ‘आंग्लभारती’ अँग्रेजी से भारतीय भाषाओं के लिए मशीनी अनुवाद प्रणाली का विकास कार्य शुरू किया गया | इसके बाद अँग्रेजी से हिंदी अनुवाद ‘आंग्लहिंदी’ (AnglaHindi) नामक सॉफ्टवेयर बनाया गया| ‘अनुभारती’ हिंदी से अँग्रेजी में अनुवाद करने के लिए इसका विकास किया गया |

अनुसारका (ANUSAARAKA) – यह परियोजना प्रो. राजीव संगल के निर्देशन में प्रारंभ हुई जो तेलुगु, कन्नड़, पंजाबी, मराठी और बंगला से हिंदी में अनुवाद प्रणाली के विकास हेतु बनाई गई है। यह पाणिनि व्याकरण के सिद्धांतों का प्रयोग कर विकसित की गई है | जो <http://ltrc.iiit.ac.in/showfile.php?filename=downloads/anu/index.htm> इस वेबसाईट पर उपलब्ध है |

b) सी-डैक पुणे (C-DAC Pune) – सी-डैक पुणे में प्रो. हेमंत दरबारी के निर्देशन में भाषा प्रौद्योगिकी से जुड़े कई महत्वपूर्ण कार्य किए जा रहे हैं जिसमें ‘मंत्रा’ (2007) नामक सॉफ्टवेयर का विकास कार्य चल रहा है जो राजभाषा के प्रशासनिक, वित्तीय, कृषि, लघु उद्योग, सूचना प्रौद्योगिकी, स्वास्थ्य, रक्षा, शिक्षा एवं बैंकिंग क्षेत्रों के से संबंधित पत्र, अधिसूचना तथा विज्ञप्तियाँ आदि को अँग्रेजी से हिंदी में अनुवाद करता है। ‘मंत्रा’ राजभाषा भारत सरकार के सभी मंत्रालयों में मानक तथा शीघ्र गति से हिंदी अनुवाद में

सहायक होगा। इसके साथ-साथ वाक् प्रौद्योगिकी से संबंधित 'प्रवाचक' (पाठ से वाक् प्रणाली) और 'वाचांतर' (वाक् से पाठ प्रणाली) पर भी अनुसंधान कार्य शुरू है। हिंदी शिक्षण के लिए 'लीला पॅकेज' 'कंप्यूटर की सहायता से प्रबोध, प्रवीण तथा प्राज्ञ स्तर की हिंदी सभी भारतीय भाषाओं के माध्यम से स्वयं सीखने के लिए राजभाषा विभाग ने कंप्यूटर प्रोग्राम (लीला हिंदी प्रबोध, लीला हिंदी प्रवीण, लीला हिंदी प्राज्ञ) तैयार करवा कर सर्व साधारण द्वारा उसका निशुल्क प्रयोग के लिए उसे राजभाषा विभाग की वेब साइट पर उपलब्ध कर दिया है।

c) सी-डैक मुंबई (C-DAC Mumbai) – मुंबई में **मात्रा (MaTra)** – अनुवाद प्रणाली का विकास 2004 में किया गया जो समाचारों, वार्षिक प्रतिवेदनों और तकनीकी पदबंधों के क्षेत्र में कार्य करने वाली यह प्रणाली अँग्रेजी पूर्वसर्गों का हिंदी परसर्गों में प्रतिचित्रण करती है। यह प्रणाली संस्था के साइट **मात्रा 2 (MaTra)** – यह अँग्रेजी-हिंदी स्वयंचलित मशीनी अनुवाद प्रणाली है जो <http://www.cdacmumbai.in/mattra/> पर उपलब्ध है।

'**Xlit**' – यह लिप्यांतरण (Transliteration) के लिए बनाया गया उपकरण है, जो अँग्रेजी के शब्दों को मराठी उर्दू, कन्नड़ और हिंदी भारतीय भाषाओं में लिप्यंतरित करता है जो <http://extensions.services.openoffice.org/project/xlithindi> इस वेबसाइट पर उपलब्ध है |

StatMT – यह सांख्यिकीय मशीनी अनुवाद(SMT) प्रणाली है जो अँग्रेजी से भारतीय भाषाओं के लिए बनाई गई है | इससे संबंधित जानकारी <http://www.cdacmumbai.in/e-ilmt>. इस वेबसाइट पर उपलब्ध है |

SuTra – यह मल्टीयुसर अनुवाद सहायक उपकरण है जो <http://sourceforge.net/projects/sutra/>. इस पर उपलब्ध है |

Rupantar – यह किसी भी लिपि में लिखे गए सामग्री को दुसरे लिपि में रूपांतरित करता है जिसका हिंदी, मराठी, तेलुगु, मलयालम, तमिल, गुरुमुखी, गुजराती, कन्नड़, उड़िया, बांग्ला इन भाषाओं के लिपि रूपांतर

के लिए उपयोग किया जा सकता है। यह उपकरण <http://www.cdacmumbai.in/rupantar> इस पर उपलब्ध है।

ChitranTran – यह किसी टेक्स्ट समग्री को प्रतिमा में रूपांतरित करने वाला लिप्यांतरण उपकरण है जो अंग्रेजी और हिंदी तथा अन्य भारतीय भाषाओं के लिए बनाया गया है।

शब्दांजली (Shabdanjali) – सी-डैक मुंबई ने शब्दांजली नामक अंग्रेजी-हिंदी द्विभाषी शब्दकोश का विकास किया है जो <http://www.cdacmumbai.in/matra/dict.jsp> इस लिंक पर उपलब्ध है।

d) IIT मुंबई – IIT मुंबई में हिंदी और मराठी के लिए शब्दतंत्र, अंग्रेजी-हिंदी शब्दकोश का विकास किया गया है।

e) IIIT हैद्राबाद – ‘शक्ति’ (शक्ति-किट, 2005) यह अनुवाद प्रणाली ‘अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैद्राबाद में विकास कार्य किया जा रहा है जिसमें अंग्रेजी से हिंदी, मराठी और तेलुगु इस भारतीय भाषाओं से संबंधित कार्य किया जा रहा है। यह अनुवाद प्रणाली वेबसाइट <http://shakti.iiit.net> पर उपलब्ध है।

f) म.गां. अं. हिंदी विश्वविद्यालय, वर्धा – वर्धा हिंदी शब्दकोश, सक्षम हिंदी वर्तनी परीक्षक (Hindi Spell Checker), समाज विज्ञान विश्वकोश आदि सॉफ्टवेयर <http://www.hindivishwa.org> इस लिंक पर उपलब्ध है। यहाँ यूनिकोड से कृतिदेव से यूनिकोड लिपि परिवर्तक भी उपलब्ध है।

g) भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (TDIL) – के वेबसाइट से 525 हिंदी फॉन्ट, फॉन्ट कोड परिवर्तक, वर्तनी संशोधक, शब्द संसाधक, कार्पोरा, ट्रांसलिटरेशन, ओ सी आर, अंग्रेजी -हिंदी शब्दकोश, हिंदी स्पेल चेकर को निशुल्कप्रयोग के लिए वेब साइट पर उपलब्ध करा दिया गया। इन्हें <http://ildc.in> से डाउनलोड किया जा सकता है। इसके साथ साथ मशीनी अनुवाद प्रणालियाँ, पार्ट ऑफ़ स्पीच टैगर, रूपिम विश्लेषक(Morphological Analyzer) भी उपलब्ध है।

TDIL

के

http://www.tdil-dc.in/components/com_mtsystem/CommonUI/homeMT.php

इस वेबसाइट पर 'अंग्ला'(AnglaMT) और 'अनुवादक्ष'(ANVADAKSH) ये दो मशीनी अनुवाद प्रणालियाँ उपलब्ध है जो अँग्रेजी, हिंदी, मराठी, उर्दू, पंजाबी, तमिल, बंगाली, मलयालम, उड़िया, तेलुगु भाषाओं के लिए है जिसका स्वास्थ्य, पर्यटन और सामान्य क्षेत्र के अनुवाद कार्य के लिए प्रयोग किया जाता है।

'अंग्ला' (AnglaMT) – अँग्रेजी-बंगाली, अँग्रेजी-मलयालम, अँग्रेजी-उर्दू और अँग्रेजी-पंजाबी भाषाओं के लिए IIT कानपूर ने डॉ. आर. एम. के सिन्हा के निर्देशन में बनाया।

'संपर्क' (SAMPARK) – भारतीय भाषाएँ हिंदी-पंजाबी, पंजाबी-हिंदी, तेलुगु-तमिल और उर्दू-हिंदी भाषा में अनुवाद करने के लिए यह अनुवाद प्रणाली उपलब्ध है जिसको IIIT हैद्राबाद ने डॉ. राजीव सेंगल के निर्देशन में विकसित किया है।

'अनुवादक्ष' (ANVADAKSH) – यह मशीनी अनुवाद प्रणाली 'अँग्रेजी से भारतीय भाषा मशीन अनुवाद प्रणाली' (E-ILMT) है। जिसमें अँग्रेजी-हिंदी, अँग्रेजी-मराठी, अँग्रेजी-बंगाली, अँग्रेजी-उड़िया और अँग्रेजी-तमिल इन भाषाओं के लिए प्रणाली उपलब्ध है। इसपर वर्तमान में आगे का विकास कार्य भारत के कई संस्थानों में चल रहा है।

'संधान' (sandhan) यह एकभाषी सर्च सिस्टम है, जो बंगाली, हिंदी, मराठी, तमिल और तेलुगु भाषाओं में पर्यटन क्षेत्र की जानकारी की खोज के लिए बनाया गया है।

h) **'अनुवादोक' (Anubadok)** – अँग्रेजी से बँगला अनुवाद प्रणाली जिसमें कुछ सिमित शब्दों का ही अनुवाद करती है। <http://bengalinux.sourceforge.net/cgi-bin/anubadok/index.pl> इस पर ऑनलाइन उपलब्ध है।

जवाहरलाल नेहरू विश्वविद्यालय, दिल्ली में 'संस्कृत अध्ययन केंद्र' द्वारा संस्कृत-हिंदी अनुवाद के लिए 'सहित' (SaHit) नामक प्रणाली का विकास किया जा रहा है। 'अनुवादक' यह अनुवाद प्रणाली सुपर इंफोसॉफ्ट प्राइवेट लिमिटेड (Super Infosoft Pvt. Ltd.), दिल्ली द्वारा श्रीमती अंजली रावचौधरी के निर्देशन में विकसित अंग्रेजी से हिंदी अनुवाद करने वाला अनुवादक 5.0 पश्च संपादन में सहायक है। इसके साथ-साथ एक अंतरराष्ट्रीय स्तर की 'सर्वभाषिक संपर्क भाषा' (UNL: Universal Networking Languages) नामक परियोजना पर भी काम किया जा रहा है जिसका उद्देश्य 'संयुक्त राष्ट्र संघ' के माध्यम से विश्व की अधिक से अधिक भाषाओं में संपर्क बढ़ाना है।

भारत में भारतीय भाषाओं के लिए कार्य करने वाले संस्थान निम्न हैं –

- मराठी और कोंकणी – IIT मुंबई और C-DAC पुणे।
- असमी और मणिपुरी – IIT गुवाहाटी।
- बांग्ला – भारतीय सांख्यिकीय संस्थान, कोलकाता।
- विदेशी भाषा (जापानी, चीनी) और संस्कृत (भाषा अभिगम प्रणाली) – जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली।
- गुजराती – एम.एस. विश्वविद्यालय, बड़ौदा।
- हिंदी और नेपाली – भारतीय प्रौद्योगिकी संस्थान, कानपुर।
- कन्नड़ और संस्कृत (संज्ञानात्मक मॉडल) – भारतीय विज्ञान संस्थान, बंगलौर।
- मलयालम – प्रगत संगणन विकास केंद्र (C-DAC) तिरुवनंतपुरम।
- उड़िया – कंप्यूटर विज्ञान और अनुप्रयोग विभाग, उत्कल विश्वविद्यालय।
- पंजाबी – थापर इंजीनियरिंग एवं प्रौद्योगिकी संस्थान, पटियाला।
- तमिल – अन्ना विश्वविद्यालय, चेन्नई।
- तेलुगु – हैदराबाद विश्वविद्यालय, हैदराबाद।
- उर्दू, सिंधी – सी-डैक, पुणे।

1. गूगल अनुवाद प्रणाली Google Translator – <http://translate.google.co.in/> – कुल 80 भाषाएँ जिसमें मराठी के साथ साथ ६ भारतीय भाषाओं में अनुवाद कार्य किया जा रहा है | अन्य ऑनलाईन मशीनी अनुवाद प्रणालियाँ –
2. बिंग अनुवाद प्रणाली (bing Translator) – कुल 44 भाषाएँ जिनमें 3 भारतीय भाषाओं में अनुवाद कार्य किया जा रहा है |
3. फ्री ट्रांसलेशन – [sld.freetranslation](http://sld.freetranslation.com/) – कुल 43 जिनमें 3 भारतीय भाषाओं में अनुवाद कार्य किया जा रहा है |
4. Babylon translator – <http://translation2.paralink.com/English-Hindi-Translator>
5. <http://www.word2word.com/free.html> (शब्दनुवाद/वर्ड-टू-वर्ड) इसपर कई भाषाई उपकरण उपलब्ध है |
6. Aktuel Translation – <http://www.aktueltranslations.com/free-online-translation/> इसपर भी विदेशी और भारतीय भाषाओं के लिए अनुवाद प्रणाली उपलब्ध है |
7. <http://www.onehourtranslation.com/free-translation/> इसपर भी कुल 110 भाषाओं में 14 भारतीय भाषाओं के लिए अनुवाद प्रणाली उपलब्ध है |

ऑनलाईन ई-कोश –

संगणकीय कोशकारिता – कोशकारिता यह एक बहुत ही क्लिष्ट और परिश्रम वाला कार्य है। परिवर्तन भाषा विशेषता के कारण समय के साथ-साथ भाषा के प्रचलित शब्दों और अर्थों में निरंतर बदलाव होता रहता है, नए शब्दों का आगमन, शब्दों का लोप आदि के कारण पुराने कोशों अर्थात् पुस्तकी कोश में समय-समय पर पुनः सुधार करना संभव नहीं है | इस लिए पुनः उन कोशों के नए संस्करण का निर्माण करना पड़ता है और यह बहुत समय लेने वाला, कठिन और श्रम वाला कार्य है | इसलिए आज कोशकारिता जैसी विषय को भी कंप्यूटर के साथ जोड़कर शब्दों में निरंतर होने वाले परिवर्तन, उसमें होने वाले लुप्त, नए

शब्दों का गढ़ना आदि को समय समय पर सुधार करने के लिए 'संगणकीय कोशकरीता' (Computational Lexicography) जैसी संकल्पना ने जन्म लिया जिससे आनेवाले समय में कम समय में शब्दों की खोज उनके अर्थ को जानने- समझने के लिए सरलता आ गई | इंटरनेट पर अँग्रेजी से भारतीय भाषाओं के लिए कई कोश ऑनलाइन उपलब्ध है जिसकी वेबसाईट नीचे दिए गए हैं –

1. वर्डनेट (शब्दतंत्र) – मराठी और हिंदी के लिए IIT मुंबई द्वारा विकसित हिंदी और मराठी वर्डनेट, हिंदी-अँग्रेजी शब्दकोश <http://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php> इस वेबसाईट पर उपलब्ध है।
2. शब्दकोश.काम – <http://www.shabdkosh.com/> इस वेबसाईट पर अँग्रेजी से हिंदी, मराठी, बांग्ला, गुजराती, कन्नड़, मलयालम, पंजाबी, तमिल, तेलुगु आदि के लिए शब्दकोश उपलब्ध है।
3. खंडबहाले शब्दकोश – <http://www.khandbahale.org/> इस वेबसाईट पर अँग्रेजी से हिंदी, मराठी, संस्कृत, बांग्ला, गुजराती, कन्नड़, मलयालम, पंजाबी, तमिल, तेलुगु, नेपाली, डोगरी, कोंकणी, बोडो, असमिया, मैथिली, संथाली, सिंधी, कश्मीरी, उर्दू, मणिपुरी, उड़िया आदि सभी भारतीय भाषाओं के लिए शब्दकोश उपलब्ध है |
4. ई-शब्दकोश. कॉम – <http://www.eshabdakosh.com/> इस पर अँग्रेजी से मराठी शब्दकोश उपलब्ध है |
5. हिंदी खोज शब्दकोश – <http://dict.hinkhoj.com/> इस पर उपलब्ध है |
6. अँग्रेजी-हिंदी के लिए – <http://hindi.changathi.com/Dictionary.aspx> इस वेबसाइट पर शब्दकोश उपलब्ध है |
7. solpro शब्दकोश – www.solpro.com पर ऑफ़लाइन शब्दकोश अँग्रेजी से हिंदी के लिए उपलब्ध है जो डाउनलोड किया जा सकता है |
8. offlinehindidictionary.com (मोबाईल अप्स) – इसपर अँग्रेजी-हिंदी-अँग्रेजी शब्दकोश उपलब्ध है |

अन्य स्रोत अँग्रेजी, हिंदी तथा अन्य भाषाओं से संबंधित उसके साहित्य, भाषाई व्याकरण आदि से संबंधित कई वेबसाइट्स उपलब्ध है जिनको हम निम्न प्रकार से देख सकते हैं –

1. 'हिंदी सबके लिए' <http://hindiforyou.blogspot.in/2009/09/blog-post.html> इस ब्लॉग पर व्याकरण, कोश, हिंदी शिक्षण पाठ्यसामग्री, हिंदी साहित्य तथा हिंदी से जुड़ी कई सामग्री उपलब्ध है |
2. 'हिंदी टेक' यह 'भाषा-प्रौद्योगिकी' से संबंधित ऑनलाइन पत्रिका है जहाँ इससे संबंधित आलेख, भाषाई उपकरण – POS टैगर, हिंदी वर्तनी जाँचक, हिंदी मानकीकरण टूल, बहुभाषिक शब्दकोश(अँग्रेजी-हिंदी-मराठी), चंकर आदि उपलब्ध हैं | जो <http://hinditech.in/> पर उपलब्ध हैं |
3. क्विक समाधान नामक इस <http://quicksamadhan.com> वेबसाइट पर हिंदी, अँग्रेजी, मराठी, तमिल, बंगाली, तेलुगु, गुजराती, पंजाबी, मलयालम, उर्दू, मैथिली, भोजपुरी, उडिया आदि भारतीय भाषाओं के व्याकरण से संबंधित जानकारी उपलब्ध है |
4. हिंदी भाषा से संबंधित हिंदी समय (<http://www.hindisamay.com/>), <http://learn-hindi-online.com/hindi-grammar/>, <http://www.hindikunj.com> , <http://hindigrammarsite.blogspot.in> आदि वेबसाइट उपलब्ध हैं |
5. अँग्रेजी भाषा व्याकरण से संबंधित <http://www.verbix.com/webverbix/English/run.html> , <http://www.englishpage.com/> आदि वेबसाइट उपलब्ध हैं |
6. मराठी भाषा से संबंधित http://mylanguages.org/learn_marathi.php, <http://ilanguages.org/marathi.php>, <http://www.marathidictionary.org/wordmeaning.php> उपलब्ध हैं |

7. संस्कृत भाषा से संबंधित <http://learnsanskrit.org/grammar>,
<http://www.sanskritebooks.org>,
<http://sanskrit.samskrutam.com/en.MainPage.ashx>, आदि वेबसाईट उपलब्ध हैं |
8. अंग्रेजी के लिए www.oxforddictionaries.com पर Online Oxford Dictionary तथा अन्य कई शब्दकोश उपलब्ध हैं |
9. सर्च इंजिन – गूगल, याहू, बिंग, अल्ट्रा विस्टा आदि सर्च इंजिन लगभग सभी भाषाओं में उपलब्ध है जिससे हम किसी भी जानकारी किसी भी भाषा में पा सकते हैं |
10. सभी भारतीय भाषाओं हेतु Google Transliteration IME Tool, Google Input Tool आदि टूल उपलब्ध हैं | जो टंकण करने के लिए बहुत सरल है | जो डाउनलोड किए जा सकते हैं | अंग्रेजी से भारतीय भाषाओं के लिए <http://utilities.webdunia.com/hindi/transliteration.html> इस वेबसाईट पर लिप्यंतरण उपकरण उपलब्ध है |

संसार में होने वाली सभी भाषाओं का ज्ञान को अर्जित करना और सभी कार्य एक साथ करना एक मनुष्य के लिए संभव नहीं है इसलिए मनुष्य ने अपने कार्य को सरल बनाने और अपने आवश्यकता के लिए यंत्रों का विकास किया होगा जिनका आवश्यकता के अनुरूप वह उपयोग कर सके | आज बदलते परिवेश और प्रौद्योगिकी के क्षेत्र में दिन-ब-दिन होने वाले साधनों के आविष्कार और विकास के कारण उसके साथ-साथ संपर्क-तकनीक और संपर्क-साधनों में भी बहुत बड़ा बदलाव आ गया है | आज समय के साथ नहीं समय के आगे रहना जरूरी है | क्योंकि आज बाजार में उत्पाद कंपनियों में लगी होड़ के कारण आज बाजार में हर दिन नई-नई तकनीक युक्त उपकरण उपलब्ध हो रहे हैं | आज का उपभोगता भी एक साथ एक जगह कम समय में कई सुविधाएँ चाहता है इसलिए इस माँग और आवश्यकता की पूर्ति के लिए नए-नए उत्पाद बाजार में आने लगे हैं | जिनका हम अपने सुविधा के अनुसार जब चाहे उपयोग कर सकते हैं | आज समाज में दिन-ब-दिन नई तकनीक की माँग हो रही है जिसके कारण इंटरनेट, मोबाइल, संगणक में नई

तकनीकों का प्रयोग किया जाने लगा है साथ ही नए-नए ऑनलाईन तथा ऑफ़लाईन सॉफ्टवेयर उपलब्ध किए जा रहे हैं।

वर्तमान में देखा जाए तो भारत और विदेशों में प्राकृतिक भाषा संसाधन के क्षेत्र में कई कार्य हुए हैं और हो भी रहे हैं। आज ‘भाषा-प्रौद्योगिकी’ के चलते भाषा और कंप्यूटर के क्षेत्र से संबंधित होने वाले छात्रों, शोधार्थियों, अध्यापकों के लिए इस क्षेत्र में काम करने के लिए बहुत बड़ा मंच उपलब्ध हो गया है। कहने का तात्पर्य यह है कि आज भाषा विज्ञान या भाषा से संबंधित होने वाले छात्रों के लिए आज के समय में रोजगार की कई संभावनाएँ हैं। आज भारत के कई विश्वविद्यालयों में भाषा विज्ञान, अनुवाद प्रौद्योगिकी, भाषा प्रौद्योगिकी, कंप्यूटेशनल भाषाविज्ञान, भाषा अभियांत्रिकी जैसे विभाग खुल गए हैं, जहाँ प्राकृतिक भाषा संसाधन से संबंधित भारतीय भाषाओं का अध्ययन और शोध कार्य भी किए जा रहे हैं। आज भाषाओं का इसी परिप्रेक्ष्य में तुलनात्मक अध्ययन करने की आवश्यकता है। जिससे इस कार्य को सरल बनाने हेतु बहुत ही मदद होगी। जिससे आनेवाले समय में निश्चित ही ‘ई-भाषाई समाज’ का निर्माण होगा और सारे समाज की ‘संप्रेषण माध्यम भाषा’ एक सी होगी जिसके कारण सभी भाषाओं को हम अपना कह सकेंगे।

संदर्भ सूची –

- Bharti, Akshar. Chaitanya, Vinit. Sangal, Rajeev (2000) Natural Language Processing: A Paninian Perspective; Prentice Hall of India, New Delhi.
- त्रिपाठी, अरिमर्दन कुमार. (2012) ‘भाषा के समकालीन संदर्भ’, साहित्य संगम, इलाहाबाद |
- मल्होत्रा, विजय कुमार. (2002) कंप्यूटर के भाषिक अनुप्रयोग, साहित्य सहकार, नई दिल्ली|
- <http://tdil.mit.gov.in/>
- http://www.cdacmumbai.in/index.php/cdacmumbai/research_publication_s/research_groups/kbcs_artificial_intelligence/products_projects/ongoing_projects/matra



- <http://matra2.blogspot.in/2012/05/multilingual-computing-at-c-dac-mumbai.html>
- <http://www.rajbhasha.nic.in/IIContent.aspx?t=techevents>
- <http://www.hinditech.in>
- <https://www.google.co.in/search?newwindow=1&site=&source=hp&q=translation&oq=translation>

Citation: झाल्टे, धनंजय विलास (2015). भारतीय भाषाओं हेतु डिजिटल संसाधन, HindiTech: A Blind Double Peer Reviewed Bilingual Web-Research Journal, 6 (8), 92-104.

URL: <https://hinditech.in/bhartiy-bhashaon-hetu-digital-sansadhan/>