

## भाषा-अभियांत्रिकी का उद्भव एवं विकास

प्रवेश कुमार द्विवेदी

[dpraveshkumar@gmail.com](mailto:dpraveshkumar@gmail.com)

महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय से भाषा-अभियांत्रिकी में पी-एच.डी.

---

आज का युग अभियांत्रिकी का युग है, जहाँ एक तरफ संगणक, मोबाइल, पी.डी.ए. आदि उत्पादों की डिजाइनिंग, निर्माण आदि कार्य किए जाते हैं, तो दूसरी तरफ उनके संचालन हेतु सॉफ्टवेयर का निर्माण किया जाता है। ये सॉफ्टवेयर अथवा हार्डवेयर किसी न किसी प्राकृतिक भाषा से संबंधित होते हैं, जिनका निर्माण भाषा-अभियांत्रिकी के अंतर्गत किए जाते हैं। प्रमुख रूप से भाषा-अभियांत्रिकी का उद्भव संगणक के लिए ऐसे सॉफ्टवेयर और हार्डवेयर का निर्माण के लिए हुआ, जो प्राकृतिक भाषा से संबंधित हों। प्राकृतिक भाषा का संसाधन पाठ तथा वाक् दो रूपों में किया जा सकता है अतः पाठ तथा वाक् संसाधन से संबंधित सॉफ्टवेयर तथा हार्डवेयर दोनों का सीधा संबंध 'भाषा-अभियांत्रिकी' से है। यदि सीधे तौर पर देखा जाय, तो अभियांत्रिकी की वह प्रक्रिया है, जिसके माध्यम से प्राकृतिक भाषा की अभिकल्पना, विकास, जाँच-परख एवं रख-रखाव किया जाता है। जैसा कि अभियांत्रिकी की अन्य विधाओं के विषय से हमें ज्ञात होता है कि वर्तमान भाषा प्रणाली का उद्भव और विकास पारंपरिक भाषावैज्ञानिक अध्ययन और संगणक विज्ञान के बुनियाद पर ही हो रहा है।

कंप्यूटर में तैयार किए गए प्रोग्राम्स के समूह को सॉफ्टवेयर कहते हैं। कंप्यूटर हार्डवेयर को संचालित करने के लिए मानव द्वारा निर्देश देने की विधि सॉफ्टवेयर के रूप में होती है। कंप्यूटर के यांत्रिक, विद्युत तथा इलेक्ट्रॉनिक भाग हार्डवेयर कहलाते हैं। यहाँ पर भाषा से संबंधित यांत्रिक, विद्युत तथा इलेक्ट्रॉनिक युक्ति

का निर्माण किया जाता है। संगणक के विकास के बाद से ही इसमें भाषिक ज्ञान को स्थापित करने के प्रयास आरंभ हो गए जो अभी तक चल रहे हैं। धीरे-धीरे इसमें सीमित सफलताएं भी मिली है और किंतु किसी भी भाषा के संसाधन में पूर्णतः सफलता अभी तक प्राप्त नहीं हो सका है। इस दौरान मोबाइल और अन्य स्वचलित मशीनों का भी तेजी से अविष्कार किया गया है। जिनमें भाषाई ज्ञान जितना अधिक होगा, उनकी उपयोगिता उतनी ही बढ़ती जाएगी। यही कारण है कि संपूर्ण विश्व में भाषा से जुड़े सॉफ्टवेयरों और अनुप्रयोग प्रणालियों का तेजी से विकास किया जा रहा है। इसके लिए पिछले पाँच दशकों में अनेक क्षेत्र उभर कर सामने आए हैं, जिनमें भाषा-अभियांत्रिकी प्रमुख है। भाषा-अभियांत्रिकी में 'प्राकृतिक भाषा संसाधन' की प्रक्रिया निरंतर चलती रहती है।

भाषा-अभियांत्रिकी को समझने के लिए हमें पहले भाषा तथा अभियांत्रिकी को अलग-अलग समझना होगा, जो इस प्रकार से है-

**भाषा** – यादृच्छिक वाक् प्रतीकों की वह व्यवस्था जिसके माध्यम से मानव समुदाय आपस में विचारों-भावों का आदान-प्रदान करते हैं, भाषा कहलाती है। कॉलिन्स अँग्रेजी शब्दकोश में भाषा को इस प्रकार परिभाषित किया है "मानव अपने विचारों, भावों आदि को बोलकर या संकेत के द्वारा अभिव्यक्ति करता है, भाषा कहलाती है।" भाषा संचार का तंत्र है जो लिखित या वाचिक होती है।

**अभियांत्रिकी** – वैज्ञानिक ज्ञान का प्रयोग करके अनुप्रयुक्त क्षेत्र में आने वाली समस्याओं का समाधान करना अथवा वैज्ञानिक सिद्धांत का प्रयोग करके मशीन, इंजन, कार आदि का डिजाइन, निर्माण कार्य और प्रारूपण करने को ही अभियांत्रिकी कहते हैं। अभियांत्रिकी विधियों के विकास का समुच्चय है। अभियांत्रिकी मानव निर्मित वस्तुओं के माध्यम से सैद्धांतिक ज्ञान को व्यवहार में लाने की विधि है। अभियांत्रिकी तकनीकी का मूर्त पक्ष है।

अतः यह कहा जा सकता है कि भाषावैज्ञानिक भाषा से संबंधित नए ज्ञान बनाने में रुचि रखते हैं जबकि भाषा-अभियांत्रिकी इच्छित लक्ष्य के अनुसार प्रणाली बनाने में रुचि रखते हैं। भाषा-अभियांत्रिकी के अंतर्गत प्रकृति और उसके सामग्री व्यवहार और भौतिक जगत का व्यवस्थित, अवलोकन, प्रयोग, माप,

नियमों का निर्माण, तथ्यों आदि को अध्ययन कर उसे मूर्त पक्ष दिया जाता है। गतिशील और बहुआयामी एक अनुसंधान के क्षेत्र में मानव भाषा के साथ परिभाषित करना एक कठिन कार्य है। भाषा-अभियांत्रिकी को कुछ वैज्ञानिकों ने इस प्रकार से परिभाषित किया है-

**गजदार (1996) के अनुसार** – “भाषा-अभियांत्रिकी कम्प्यूटर विज्ञान का एक ऐसा विषय है जिसमें ऐसे Intelligent Computational artifacts बनाए जाते हैं, जो संगणक भाषावैज्ञानिक के अलावा अन्य लोगों के लिए भी उपयोगी हो इसकी उपयोगिता की जाँच बाजार माँग के उपर निर्भर करता है।”

**थोम्पसन (1985) के अनुसार-** “भाषा-अभियांत्रिकी, प्राकृतिक भाषा संसाधन का एक एप्लीकेशन है जिसमें ऐसे कम्प्यूटर सिस्टम का निर्माण करना है जो भाषा को संसाधित कर सके और भाषा से संबंधित कुछ कार्य किया जा सके।” (Thompson, 1985)

**जैकब (1992) के अनुसार-** “भाषा-अभियांत्रिकी के द्वारा बनाए गए प्रणाली को प्रभावशाली, तेज और उपयोगी होना चाहिए।”

“भाषा-अभियांत्रिकी का संबंध कम्प्यूटर आधारित पाठ तथा वाक् संसाधन से है।”

अतः भाषा-अभियांत्रिकी को पारिभाषित करते हुए यह कहा जा सकता है कि “भाषा-अभियांत्रिकी, संगणक अभियांत्रिकी की वह शाखा है जिसमें प्रणाली में कृत्रिम बुद्धि को प्रदान कर इसतरह तैयार किया जाता है कि वह मानवीय भाषा को समझ कर ठीक उसी प्रकार व्यवहार करे जिस प्रकार मनुष्य भाषा को समझ कर कार्य करता है।” भाषा-अभियांत्रिकी एक ऐसा अनुशासन है जिसके अंतर्गत मानवीय भाषा से संबंधित कार्यों को संसाधित किया जाता है एवं जिसकी संसाधन प्रक्रिया और प्राप्त परिणाम दोनों उम्मीद के मुताबिक और औसत दर्जे का हो। प्रासंगिक वैज्ञानिक परिणाम और अभ्यास की संरचना दोनों ही साहित्य के क्षेत्र से संबंधित हैं। भाषा-अभियांत्रिकी के अंतर्गत इलेक्ट्रॉनिक उत्पादों के साथ साथ भाषिक अनुप्रयोग वाले प्रणाली को सहजतम रूप से तैयार किए जा सके। अर्थात् जिस तरह से अन्य अभियांत्रिकी में किसी न किसी उत्पाद को तैयार किया जाता है ठीक उसी प्रकार यहाँ भी भौतिक उत्पादन के साथ साथ

भाषिक अनुप्रयोग तैयार किया जाना चाहिए जिसका प्रयोग संगणक भाषाविद के साथ साथ अन्य उपयोगकर्ता इसका उपयोग कर सकें।

भाषा-अभियांत्रिकी को जब ऐतिहासिक परिप्रेक्ष में देखा जाए, तो इसका इतिहास बहुत पुराना नहीं है। इसके इतिहास निम्न बिंदुओं के द्वारा देखा जा सकता है-

- प्रथम स्वचालित कम्प्यूटर का विकास सन् 1946 में इनआक के रूप में हुआ। इसके बाद से विभिन्न क्षेत्रों में इसका उपयोग किया जाने लगा।
- निर्माण और स्वरूप के दृष्टि से धीरे-धीरे इसमें आवश्यकता अनुसार अनेक परिवर्तन किए गए हैं।
- भाषायी दृष्टि से इसका व्यवहार व्यावसायिक मशीनी अनुवाद प्रणाली के क्षेत्र में था।
- सन् 1956 से कृत्रिम बुद्धि के क्षेत्र में महत्वपूर्ण कार्यों की शुरुआत हुई। इसके बाद से कम्प्यूटर में भाषा और बुद्धि से जुड़े, सॉफ्टवेयर, प्रोग्राम्स, हार्डवेयर के विकास का कार्य तेजी से होने लगा।
- भाषा प्रौद्योगिकी और कम्प्यूटेशनल लिंग्विस्टिक्स जैसे विषयों में प्राकृतिक भाषा संसाधन से जुड़े सैद्धांतिक और अनुप्रयुक्त कार्य पिछले 30 से 40 वर्षों में बृहद स्तर पर किये गये हैं, इसके साथ ही कम्प्यूटर में भाषा से संबंधित हार्डवेयर और सॉफ्टवेयर आदि को डिजाइनिंग, निर्माण, सुधार, प्रोग्रामिंग आदि की आवश्यकताओं को देखते हुये भाषा-अभियांत्रिकी का उद्भव हुआ।
- भाषा-अभियांत्रिकी पद को MITKOV (1995) से जोड़कर देखा जाता है। इसका प्रस्ताव NAGAO द्वारा COLLING में दिया गया। इस प्रोग्राम के संबंध में 1996 में EUROPEAN COMMISSION (EC) ने कहा कि “Language Engineering has successfully promoted a shift from language form research activities to more immediately feasible and industrially relevant RTD themes”.

- इधर पिछले कुछ वर्षों में ARPA(Advanced Research Project Agency ) ने भी एक डाटा मैनेजमेंट के लिए स्टोरेज मानक ए.पी.आई. से युक्त सक्षम शोध सॉफ्टवेयर की आवश्यकता पर बल दिया है।

संक्षेप में, वर्तमान में भाषा-अभियांत्रिकी से जुड़े सॉफ्टवेयर मुख्यतः निम्नलिखित विधियों पर कार्य कर रहे हैं-

- MARKOV-MODELS
- BRILL-STYLE-TRANSDUCERS
- CHART-PARSING

जब 1990 के दशक में देखा गया कि “प्राकृतिक भाषा संसाधन के समुदाय एवं संस्कृति पर वाक्-अभियांत्रिकों एवं अभियांत्रिकी उन्मुख दृष्टिकोण दोनों का आक्रमण हुआ” (Wilks 1996) तब इनके नियमों में जादुई रूप में परिवर्तन आया। प्राकृतिक भाषा संसाधन के शोध में पिछले दस सालों में क्रांति ला दिया गया। जिससे भाषा को केन्द्र-बिंदु मानकर संगणक प्रणाली में विभिन्न दृष्टिकोणों से व्यवहार में लाने की कोशिश की जाने लगी। भाषा-अभियांत्रिकी के भाषा की जटिलता एवं इसकी उपयोगिता को मद्दे नजर रखते हुए इसकी कुछ विशेषताएँ निर्धारित की गईं, जो इस प्रकार हैं-

- भाषा-अभियांत्रिकी में ऐसे प्रणालियों का निर्माण किया जाय, जिनकी प्रयोगिक उपयोगिता व्यापक स्तर पर ज्यादा हो।
- भाषा-अभियांत्रिकी के प्रणाली के कार्य करनी की क्षमता बड़े स्तर के मापने पर भी गुणात्मक रूप से ज्यादा हो।
- भाषा-अभियांत्रिकी के प्रणाली का मुख्य उद्देश्य उत्पादों को बनाने के लिए किसी भी तकनीकी का प्रयोग किया जा सकता है चाहे वह भाषा वैज्ञानिकों के अनुरूप सही हो अथवा न हो।

- इसके अंतर्गत बनाए गए सॉफ्टवेयर अथवा हार्डवेयर अनुप्रयोग उपयोग की दृष्टि से सभी गुण उपलब्ध होने चाहिए।
- भाषा-अभियांत्रिकी में कार्य करने के लिए विभिन्न विषय विशेषज्ञ हो सकते हैं, जो अपने अलग विषय के विशेषज्ञ हो।

भाषा-अभियांत्रिकी पूर्णरूपेण भाषा पर केंद्रित है, जिसमें अभियांत्रिकी का प्रयोग करके विलुप्त हो रही भाषाओं का पुनरुद्धार करना, भाषा का शुद्धीकरण करना, भाषा के बिगड़ते स्वरूप का सुधार करना, भाषा की विभिन्नता को देखते हुए एक रूपता प्रदान करने हेतु मानकीकरण करना, विभिन्न भाषाओं का विस्तार करना, अभियांत्रिकी का प्रयोग करते हुए बहुभाषी संचार हेतु सक्षम बनाना, भाषा संसाधित कार्य शब्दकोश, थिसॉरस और कॉर्पोरा का विकास आदि समाहित हैं। भाषा-अभियांत्रिकी के दूसरे संसाधन जिससे भाषा को सारभौमिक बनाने आवश्यकता होती है जिसके अंतर्गत आने वाले प्रमुख क्षेत्र जैसे: भाषा का ISO (International organization for standardization), स्क्रिप्ट कोड (ISO 639-2), कैरेक्टर सेट (ISO 10646-1/UNICODE), विभिन्न टेक्स्ट मार्कअप, फ़ाइल रूपान्तरण मानक, कई ऐसे जो कार्य जो विषय विशेष पर आधारित MARTIF (ISO 12200) हेतु सजीएमएल (SGML) Document Type Declarations (DTDs), ओ.यल.आइ.एफ, टी.एम.एक्स., जनरेटर, इंटरवल (INTERVAL), दस्तावेज़ी निर्माण एवं प्रबंधन के अंतर्गत शब्द जांचक, व्याकरण एवं शैली जांचक के साथ-साथ उत्कृष्ट कोश संसाधनों का निर्माण आदि हैं जो किसी भी भाषा के विकास एवं फैलाव हेतु अतिआवश्यक होते हैं। यहाँ पर यह भी आवश्यक हो जाता है कि ज्ञान के आधार पर भाषा की दीवार को हटाकर विभिन्न भाषाओं में भाषिक संसाधन को पहुंचाने का कार्य करे।

भाषा-अभियांत्रिकी प्रणाली के निर्माण कार्य एवं विस्तारित सीमा को भाषा प्रकार्य के द्वारा तय किया जा सकता है, जिसमें निम्न चरण समाहित हैं-

- किसी भी भाषिक प्रणाली का निर्माण करते वक्त सर्वप्रथम इनपुट के तौर में पाठ, वाक, चित्र आदि विद्युतीय भाषिक सामग्री को प्रदान किया जाता है जिसे मशीन संसाधित कर सके।

- भाषा-अभियांत्रिकी प्रणाली का निर्माण करते वक्त इच्छित आउटपुट प्राप्त करने हेतु एक निश्चित संसाधन की प्रक्रिया आवश्यक होती है जिसके अंतर्गत भाषा का संश्लेषण एवं विश्लेषण किया जाता है।
- उक्त प्रणाली में एक निश्चित संसाधन के पश्चात भाषा को संश्लेषित कर वाक्/पाठ को रूप में आउटपुट प्राप्त किया जाता है, जो विद्युतीय माध्यम में होता है।

भाषा-अभियांत्रिकी के अनुप्रयोग क्षेत्र काफी व्यापक है जिसे सीमित कर पाना मुश्किल होगा क्योंकि संगणक की परिकल्पना ही बिना भाषा के पूरी नहीं हो सकती। एक तरफ कुछ सारभौमिक अनुप्रयोग क्षेत्र जैसे मशीनी अनुवाद प्रणाली, वाक्-से-वाक् प्रणाली, वाक्-से-पाठ प्रणाली, पाठ-से-वाक् प्रणाली, हस्तलिखित पहचानक, प्रकाशिक अक्षर पहचानक, भाषा शिक्षण प्रणाली, ऑटोलॉजिकल प्रणाली, पाठ सारांसीकरण, प्रश्नोत्तर प्रणाली, खोज इंजन, विशेषज्ञ प्रणाली आदि है तो वही दूसरी तरफ किसी भी भाषा का संसाधन आपरेटिंग सिस्टम (किसी भी नए भाषिक सॉफ्टवेयर को संचालित करने हेतु), प्रोग्रामिंग भाषा (किसी भी नए भाषिक सॉफ्टवेयर के निर्माण हेतु) एवं ऐसे हार्ड वेयर (भाषिक सॉफ्टवेयर एवं हार्डवेयर के मध्य सामंजस्य स्थापित करने हेतु) के बिना संभव नहीं है, जिनका निर्माण कार्य किसी भी भाषा को सारभौमिक बनाने हेतु भाषा-अभियांत्रिकी में आवश्यक हो जाता है। जैसा की उपरोक्त से स्पष्ट होता है की भाषा को संरक्षित एवं संवर्धित करने हेतु एवं उपयोगकर्ता को उसकी मातृभाषा में सॉफ्टवेयर उपलब्ध करना है तो संगणक एक बेहतर माध्यम है और संगणक उक्त प्रणालियों के बगैर कोई भी कार्य को सम्पन्न कर पाने में असक्षम है।

**निष्कर्षतः** यह कहा जा सकता है कि भाषा-अभियांत्रिकी ऐसे साफ्टवेयर अभियांत्रिकी से संबद्ध है, जिसमें मानव भाषा को संसाधित करने संबंधी कार्य एवं साफ्टवेयर का निर्माण किया जाता है। यहां रचनात्मक प्रक्रिया एवं आउटपुट दोनों ही परिमित एवं प्रतिपाद्य होते हैं जिसके क्षेत्र का साहित्य प्रणाली एवं संदर्भित वैज्ञानिक परिणाम दोनों पर आधारित होता है। उपर्युक्त में भाषा-अभियांत्रिकी का इतिहास एवं विशेषताओं से जो नई आवश्यकताएं एवं क्षेत्र, अपनी प्राथमिकताओं एवं उपयोगिताओं के साथ उभकर



सामने आ रहे हैं, उनकी महत्ता को समझना एवं पुनःप्रयोग, शक्तिशाली, क्षमतावान और उत्पादन क्षमता को बढ़ाकर बड़े पैमाने में विस्तार करने हेतु महत्त्वपूर्ण साबित होगा।

## संदर्भसूची

- Cunningham, Hamish. 1998. "A Definition and short history of language engineering". Institute of Language, speech and hearing and Department of computer science, University of Sheffield, Sheffield, UK.
- Wright, Sue Ellen. "Trends in Language Engineering".

**Citation:** द्विवेदी, प्रवेश कुमार (2019). भाषा-अभियांत्रिकी का उद्भव एवं विकास, HindiTech: A Blind Double Peer Reviewed Bilingual Web-Research Journal, 10 (3), 27-34. URL: <https://hinditech.in/bhasha-abhiyantriki-ka-udbhav-evan-vikas/>