മലയാളം ഗവേഷണ ലേഖനങ്ങളുടെ നിർമ്മാണം

രചയിതാവ് [1], രചയിതാവ് [2],രചയിതാവ് [3]

[1] ഭാഷാശാസ്ത്ര വിഭാഗം, തുഞ്ചത്തെഴുത്തച്ഛൻ മലയാള സർവകലാശാല , തിരൂർ

[2] മലയാള വിഭാഗം, കേരള യൂണിവേഴ്സിറ്റി , തിരുവനന്തപുരം

[3] മലയാള വിഭാഗം, കേരള യൂണിവേഴ്സിറ്റി , തിരുവനന്തപുരം

ലേഖന സംഗ്രഹം

ഭാഷകളുടെ താരതമ്യാത്മക വിശകലനങ്ങളിൽ നിന്നാരംഭിച്ച് അവയുടെ ആന്തരിക ഘടനകളുടെയും സവിശേഷതകളുടെയും അന്വേഷണ തലങ്ങളിലേക്കെത്തിയ ശാസ്ത്രീയഭാഷാപഠന മേഖലയാണ് ഭാഷാശാസ്ത്രം. ഭാഷയുടെ വിവിധപ്രയോഗങ്ങളെയും പ്രയോഗവൈവിധ്യങ്ങളെയും അറിയുന്നതിനും വിശദീകരിക്കുന്നതിനുമായി വ്യത്യസ്ത വൈജ്ഞാനിക മേഖലകളുമായി ബന്ധപ്പെട്ടുകൊണ്ടുള്ള പഠനങ്ങൾ ഭാഷാശാസ്ത്രത്തിന്റെ സാധ്യത വർദ്ധിപ്പിച്ചു. സാമൂഹികഭാഷാശാസ്ത്രം, നരവംശഭാഷാശാസ്ത്രം, മനോവിജ്ഞാനീയഭാഷാശാസ്ത്രം തുടങ്ങിയവ ഇത്തരത്തിൽ വളർന്നു വന്ന വിജ്ഞാനശാഖകളാണ്. കമ്പ്യൂട്ടറധിഷ്ഠിത ഭാഷാപഗ്രഥനം, കമ്പ്യൂട്ടറധിഷ്ഠിത ഭാഷോപകരണങ്ങളുടെ നിർമാണം എന്നീ തലങ്ങളിലൂന്നികൊണ്ട് കമ്പ്യൂട്ടർ സയൻസ് എന്ന വൈജ്ഞാനിക മേഖലയുമായി ബന്ധപ്പെട്ട സജീവമായ ഒരു അന്തർവൈജ്ഞാനിക ഭാഷാശാസ്ത്ര മേഖലയാണ് കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രം ((Computational Linguistics)[1]. മനുഷ്യന്റെ സവിശേഷസിദ്ധികളെ യന്ത്രഗ്രാഹ്യമാക്കുകയും അവയിൽ നിന്നും കൃത്രിമമായി മനുഷ്യപ്രവർത്തനങ്ങൾ സാധ്യമാക്കുകയും ചെയ്യുന്ന യന്ത്രമനുഷ്യ (Robotic) പ്രവർത്തനങ്ങളായ കൃത്രിമബുദ്ധിയുടെ (Artificial Intelligence) മേഖലയിൽ വരുന്നവയാണ് കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രവും. മനുഷ്യഭാഷയുടെ (Natural Language) ഘടനയെയും സ്വഭാവത്തെയും കുറിച്ചുള്ള അറിവ് കമ്പ്യൂട്ടറിന് ഗ്രഹിക്കാവുന്ന കൃത്രിമഭാഷയിലൂടെ (Artificial Language) ധരിപ്പിക്കുകയും അവയിൽ നിന്ന് യന്ത്രാധിഷ്ഠിതഭാഷോപകരണങ്ങളുടെ നിർമാണവുമാണ് കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രത്തിന്റെ ഗവേഷണതലം.

മൂലപദങ്ങൾ

യന്ത്രാധിഷ്ഠിതഭാഷോപകരണം, കമ്പ്യൂട്ടറധിഷ്ഠിത ഭാഷാപഗ്രഥനം, കൃത്രിമബുദ്ധി, കൃത്രിമഭാഷ

ആമുഖം

കമ്പ്യൂട്ടറിന് നൽകുന്ന പാഠ/ശബ്ദത്തിന്റെ സ്വന-സ്വനിമ (Phonetic-Phonological), രൂപിമ (Morphological), വാക്യ (Syntactic), അർഥ (Semantic), വ്യവഹാര (Discourse), പ്രകരണ (Pragmatics) ഘടകങ്ങൾ വ്യവച്ഛേദിച്ച് തിരിച്ചറിയാനും അവയിൽ നിന്നും ഉപഭോക്താവ് ആവശ്യപ്പെടുന്നവ നിർദേശാനുസരണം തിരിച്ചു നൽകാനുമുള്ള യന്ത്രഗ്രാഹ്യത വർദ്ധിപ്പിക്കുകയാണ് സ്വാഭാവികഭാഷാസംസ്ക്കരണത്തിലൂടെ ചെയ്യുന്നത്. ഇത്തരത്തിൽ സ്വാഭാവികഭാഷയുടെ പദതലം മുതൽ വ്യവഹാരതലം വരെ സ്വാഭാവികഭാഷാസംസ്ക്കരണത്തിന്റെ അപഗ്രഥനപരിധിയിൽ വരുന്നു. സൈദ്ധാന്തികമായി നിരവധി മാതൃകകൾ ഈ മേഖലയിൽ അവതരിപ്പിക്കുന്നുണ്ട് [2]. എങ്കിലും പ്രായോഗികമായി ഉച്ചാരണവൈകല്യങ്ങൾ, വ്യക്തിഭാഷണവൈവിധ്യങ്ങൾ എന്നിവ സ്വന-സ്വനിമാപഗ്രഥനത്തിലും പദങ്ങളുടെ സാന്ദർഭിക അർഥവ്യത്യാസം, അർഥ-വ്യവഹാര രൂപങ്ങളിലെ സാമൂഹിക-സാംസ്ക്കാരിക ഘടകങ്ങളുടെ സ്വാധീനം തുടങ്ങിയവ സ്വാഭാവികഭാഷാസംസ്ക്കരണത്തിന്റെ വെല്ലുവിളികളാണ്. കമ്പ്യൂട്ടർ വിദഗ്ധരും ഭാഷാശാസ്ത്രജ്ഞരും ഏകീകരണത്തോടുകൂടി പ്രവർത്തിക്കേണ്ട മേഖലയാണിത്. ഇംഗ്ലീഷ്, സ്പാനിഷ്, അറബിക്ക് പോലുള്ള ലോകഭാഷകളിലും ഹിന്ദി, ഉറുദു, തമിഴ് തുടങ്ങിയ ഇന്ത്യൻ ഭാഷകളിലും കമ്പ്യൂട്ടർഭാഷാശാസ്ത്ര മേഖലയിൽ സജീവമായ പഠനങ്ങൾ നടന്നുകൊണ്ടിരിക്കുകയാണ്. മലയാളത്തിന്റെ സാഹചര്യത്തിൽ കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രം ഗൗരവതരമായൊരു അക്കാദമിക മേഖലയായിപ്പോലും ഇപ്പോഴും പരിഗണിച്ചു കാണുന്നില്ല [3]. ഭാഷാശാസ്ത്രത്തിന്റെ നവീന ശാസ്ത്രസമീപനങ്ങളെ സംബന്ധിക്കുന്ന ഈ ചർച്ചയിൽ കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രം, സ്വാഭാവികഭാഷാസംസ്കരണം എന്ന അന്തർവൈജ്ഞാനിക മേഖലയെ സാമാന്യമായി പരിചയപ്പെടുത്തുകയും ഈ പഠനമേഖലയുടെ പ്രാധാന്യവും സാധ്യതയും അവതരിപ്പിക്കുക എന്നതുമാണ് ഈ ലേഖനത്തിലൂടെ ഉദ്ദേശിക്കുന്നത്.

വിഭാഗം-1

ഭാഷയുടെ വിവിധപ്രയോഗങ്ങളെയും പ്രയോഗവൈവിധ്യങ്ങളെയും അറിയുന്നതിനും വിശദീകരിക്കുന്നതിനുമായി വ്യത്യസ്ത വൈജ്ഞാനിക മേഖലകളുമായി ബന്ധപ്പെട്ടുകൊണ്ടുള്ള പഠനങ്ങൾ ഭാഷാശാസ്ത്രത്തിന്റെ സാധ്യത വർദ്ധിപ്പിച്ചു. സാമൂഹികഭാഷാശാസ്ത്രം, നരവംശഭാഷാശാസ്ത്രം, മനോവിജ്ഞാനീയഭാഷാശാസ്ത്രം തുടങ്ങിയവ ഇത്തരത്തിൽ വളർന്നു വന്ന വിജ്ഞാനശാഖകളാണ് [4]. കമ്പ്യൂട്ടറധിഷ്ഠിത ഭാഷാപഗ്രഥനം,

കമ്പ്യൂട്ടറധിഷ്ഠിത ഭാഷോപകരണങ്ങളുടെ നിർമാണം എന്നീ തലങ്ങളിലൂന്നികൊണ്ട് കമ്പ്യൂട്ടർ സയൻസ് എന്ന വൈജ്ഞാനിക മേഖലയുമായി ബന്ധപ്പെട്ട സജീവമായ ഒരു അന്തർവൈജ്ഞാനിക ഭാഷാശാസ്ത്ര മേഖലയാണ് കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രം ((Computational Linguistics). മനുഷ്യന്റെ സവിശേഷസിദ്ധികളെ യന്ത്രഗ്രാഹ്യമാക്കുകയും അവയിൽ നിന്നും കൃത്രിമമായി മനുഷ്യപ്രവർത്തനങ്ങൾ സാധ്യമാക്കുകയും ചെയ്യുന്ന യന്ത്രമനുഷ്യ (Robotic) പ്രവർത്തനങ്ങളായ കൃത്രിമബുദ്ധിയുടെ (Artificial Intelligence) മേഖലയിൽ വരുന്നവയാണ് കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രവും.

ഉപവിഭാഗം

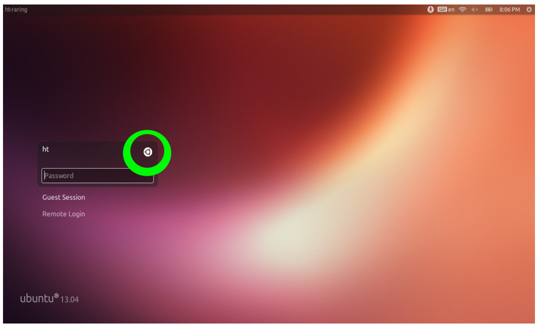
മനുഷ്യഭാഷയുടെ (Natural Language) ഘടനയെയും സ്വഭാവത്തെയും കുറിച്ചുള്ള അറിവ് കമ്പ്യൂട്ടറിന് ഗ്രഹിക്കാവുന്ന കൃത്രിമഭാഷയിലൂടെ (Artificial Language) ധരിപ്പിക്കുകയും അവയിൽ നിന്ന് യന്ത്രാധിഷ്ഠിതഭാഷോപകരണങ്ങളുടെ നിർമാണവുമാണ് കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രത്തിന്റെ ഗവേഷണതലം. ഇത്തരത്തിൽ മനുഷ്യഭാഷയെ യന്ത്രഗ്രാഹ്യമാക്കുന്ന പ്രക്രിയയെ സാങ്കേതികമായി സ്വാഭാവിക ഭാഷാസംസ്ക്കരണം (Natural Language Processing) എന്നു പറയുന്നു. 'കൃത്രിമബുദ്ധിയുടെ ഭാഗമായി വികസിച്ചുവന്ന കമ്പ്യൂട്ടറധിഷ്ഠിത സ്വാഭാവികഭാഷാപഗ്രഥനമാണ് സ്വാഭാവികഭാഷാസംസ്ക്കരണം' (Abhimanyu, Abhinav, Chandresh 2013:131). കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രത്തിന്റെ അടിസ്ഥാന വിഷയമാണിത്. മനുഷ്യഭാഷയിലെ മുഴുവൻ ഭാഷാംശങ്ങളും കമ്പ്യൂട്ടറിന് കൃത്യമായി മനസ്സിലാക്കിക്കൊടുക്കലാണ് സ്വാഭാവികഭാഷാസംസ്കരണത്തിന്റെ ധർമം.

വിഭാഗം-2 - യൂണിറ്റി ഡെസ്ക് ടോപ്പ്

നോട്ടിലസ് ഫയൽ മാനേജർ, ഫയർ ഫോക്സ് വെബ്ബ്രൗസർ, കോമ്പിസ് വിൻഡോ മാനേജർ, യൂണിറ്റി ഡോക്ക് തുടങ്ങിയ ആപ്ലിക്കേഷനുകളാണ് ഇതിൽ ഉൾപ്പെടുത്തിയിരിക്കുന്നത്. ഗ്നോം ഡെസ്ക്ടോപ്പിന്റെ ഒരു വകഭേദമാണിത് കോമ്പോസിറ്റിങ്ങ് എന്ന സാങ്കേതിക വിദ്യ പ്രയോജനപ്പെടുത്തിക്കൊണ്ടാണ് കോമ്പിസ് പ്രവർത്തിക്കുന്നത്. വളരെ മെച്ചപ്പെട്ട വിഷ്വൽ ഇഫക്ടുകൾ നൽകാൻ കോംമ്പിസിന് സാധിക്കും. കോംമ്പിസ് ഉപയോഗിക്കുമ്പോൾ ശരിയായ ഗ്രാഫിക്സ് ഡ്രൈവർ ഇല്ലെങ്കിൽ കമ്പ്യൂട്ടറിന്റെ പ്രവർത്തന വേഗം വല്ലാതെ കുറഞ്ഞുപോയതായി തോന്നാം. കമ്പ്യൂട്ടറിലെ പ്രോസെസ്സറിന്റെ ശേഷിയുടെ നല്ലൊരു ഭാഗം കോംമ്പിസ് ഉപയോഗിക്കുന്നതുകൊണ്ടാണിത് സംഭവിക്കുന്നത്. എടുത്തു പറയാവുന്ന ചില പ്രത്യേകതകൾ താഴെ പറയുന്നു.

1. ഇടതുവശത്തേക്ക് മാറിയിരിക്കുന്ന ക്ലോസ്, മിനിമൈസ് ബട്ടണുകൾ. ആപ്ലിക്കേഷൻ തുറക്കുന്ന മെനു കമ്പ്യൂട്ടറിന്റെ ഇടതു ഭാഗത്ത് ആണ്. അതുകൊണ്ട് ആപ്ലിക്കേഷൻ അടയ്ക്കുന്നതും ആ ഭാഗത്ത് ആണ് നല്ലത് എന്നൊരു കാഴ്ചപ്പാടുണ്ട്.
2. ആപ്ലിക്കേഷനുകളിലെ മെനു ബാർ ഏറ്റവും മുകളിലേക്ക് മാറിയിരിക്കുന്നു. മെനു ബാർ സ്ഥിരമായി ഒരിടത്ത് തന്നെ നിൽക്കുന്നതാണ് കൂടുതൽ നല്ലത് എന്നൊരു വാദം ഉണ്ട്. പ്രത്യേകിച്ച് ലിബ്രെ ഓഫീസ് പോലെ വലിയ മെനു ഉള്ള ആപ്ലിക്കേഷനുകളിൽ.
3. ആമസോൺ ആപ്പ് ഉൾപ്പെടുത്തി. ഇതുവഴി കമ്പ്യൂട്ടറിലെ എല്ലാ പ്രവർത്തനങ്ങളുടേയും ചുരുക്കം ആമസോൺ സർവറിലേക്ക് അയക്കാനും, അതുവഴി ഉബുണ്ടു ഉപയോഗിക്കുന്നവർക്ക് മെച്ചപ്പെട്ട ആമസോൺ ഷോപ്പിങ്ങ് അനുഭവം നൽകാൻ സാധിക്കും എന്ന് കനോനിക്കൽ ലിമിറ്റഡ് അവകാശപ്പെടുന്നു.
4. പരമ്പരാഗത പാനലിന് പകരം ഡോക്ക് ഗണത്തിൽ വരുന്ന പാനൽ ഉപയോഗിച്ചാണ് അപ്ലിക്കേഷൻ മെനുവും മിനിമൈസ് ചെയ്യുന്ന ആപ്ലിക്കേഷനുകളും കൈകാര്യം ചെയ്യുന്നത്.

ഗ്നൂലിനക്സ് സിസ്റ്റം എല്ലാം ഒരേ സമയം ഒന്നിൽ കൂടുതൽ ആളുകൾക്ക് ഉപയോഗിക്കുന്നതിന് വേണ്ടി രൂപകല്പന ചെയ്തതാണ്. ഉപയോക്താവ്, ഭാഷ, ഡെസ്ക്ടോപ്പ് എൻവിയോൺമെന്റ് തുടങ്ങിയ എല്ലാ കാര്യങ്ങളും തിരഞ്ഞെടുക്കാൻ സഹായിക്കുന്ന സംവിധാനമാണ് ലോഗിൻ മാനേജർ. ഉബുണ്ടുവിൽ lightdm എന്ന ഡിസ്പ്ലേ മാനേജർ ആണ് ലോഗിൻ സമയത്ത് ഉപയോഗപ്പെടുത്തുന്നത്.



ചിത്രം 1 : ലോഗിൻ മാനേജർ

എട്ട് തരം ഡെസ്ക് ടോപ്പ് എൻവിയോൺമെന്റുകളാണ് സാധാരണയായി ഉപയോഗിച്ച് വരുന്നത്. നിങ്ങളുടെ കമ്പ്യൂട്ടറിൽ ഒന്നിൽ കൂടുതൽ എൻവിയോൺമെന്റ് ഉണ്ടെങ്കിൽ ചിത്രത്തിലെ പച്ച വട്ടത്തിൽ ഉള്ളതുപോലെ ഒരു ചിത്രം ഉണ്ടാവും. അത് അമർത്തിയാൽ വേണ്ട എൻവിയോൺമെന്റ് തിരഞ്ഞെടുക്കാം. നിങ്ങളുടെ കമ്പ്യൂട്ടറിൽ ഒരേയൊരു ഡെസ്ക്ടോപ്പ് എൻവിയോൺമെന്റാണുള്ളതെങ്കിൽ ഈ ബട്ടൺ ഉണ്ടായിരിക്കുകയില്ല.

ഉപസംഹാരം

കമ്പ്യൂട്ടറിന് നൽകുന്ന പാഠ/ശബ്ദത്തിന്റെ സ്വന-സ്വനിമ (Phonetic-Phonological), രൂപിമ (Morphological), വാക്യ (Syntactic), അർഥ (Semantic), വ്യവഹാര (Discourse), പ്രകരണ (Pragmatics) ഘടകങ്ങൾ വ്യവച്ഛേദിച്ച് തിരിച്ചറിയാനും അവയിൽ നിന്നും ഉപഭോക്താവ് ആവശ്യപ്പെടുന്നവ നിർദേശാനുസരണം തിരിച്ച് നൽകാനുമുള്ള യന്ത്രഗ്രാഹ്യത വർദ്ധിപ്പിക്കുകയാണ് സ്വാഭാവികഭാഷാസംസ്ക്കരണത്തിലൂടെ ചെയ്യുന്നത്. ഇത്തരത്തിൽ സ്വാഭാവികഭാഷയുടെ പദതലം മുതൽ വ്യവഹാരതലം വരെ സ്വാഭാവികഭാഷാസംസ്ക്കരണത്തിന്റെ അപഗ്രഥനപരിധിയിൽ വരുന്നു. സൈദ്ധാന്തികമായി നിരവധി മാതൃകകൾ ഈ മേഖലയിൽ അവതരിപ്പിക്കുന്നുണ്ട്. എങ്കിലും പ്രായോഗികമായി ഉച്ചാരണവൈകല്യങ്ങൾ, വ്യക്തിഭാഷണവൈവിധ്യങ്ങൾ എന്നിവ സ്വന-സ്വനിമാപഗ്രഥനത്തിലും പദങ്ങളുടെ സാന്ദർഭിക അർഥവ്യത്യാസം, അർഥ-വ്യവഹാര രൂപങ്ങളിലെ സാമൂഹിക-സാംസ്ക്കാരിക ഘടകങ്ങളുടെ സ്വാധീനം തുടങ്ങിയവ സ്വാഭാവികഭാഷാസംസ്ക്കരണത്തിന്റെ

വെല്ലുവിളികളാണ്. കമ്പ്യൂട്ടർ വിദഗ്ധരും ഭാഷാശാസ്ത്രജ്ഞരും ഏകീകരണത്തോടുകൂടി പ്രവർത്തിക്കേണ്ട മേഖലയാണിത്. ഇംഗ്ലീഷ്, സ്പാനിഷ്, അറബിക്ക് പോലുള്ള ലോകഭാഷകളിലും ഹിന്ദി, ഉറുദു, തമിഴ് തുടങ്ങിയ ഇന്ത്യൻ ഭാഷകളിലും കമ്പ്യൂട്ടർഭാഷാശാസ്ത്ര മേഖലയിൽ സജീവമായ പഠനങ്ങൾ നടന്നുകൊണ്ടിരിക്കുകയാണ്. മലയാളത്തിന്റെ സാഹചര്യത്തിൽ കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രം ഗൗരവതരമായൊരു അക്കാദമിക മേഖലയായിപ്പോലും ഇപ്പോഴും പരിഗണിച്ചു കാണുന്നില്ല. ഭാഷാശാസ്ത്രത്തിന്റെ നവീന ശാസ്ത്രസമീപനങ്ങളെ സംബന്ധിക്കുന്ന ഈ ചർച്ചയിൽ കമ്പ്യൂട്ടർഭാഷാശാസ്ത്രം, സ്വാഭാവികഭാഷാസംസ്കരണം എന്ന അന്തർവൈജ്ഞാനിക മേഖലയെ സാമാന്യമായി പരിചയപ്പെടുത്തുകയും ഈ പഠനമേഖലയുടെ പ്രാധാന്യവും സാധ്യതയും അവതരിപ്പിക്കുക എന്നതുമാണ് ഈ ലേഖനത്തിലൂടെ ഉദ്ദേശിക്കുന്നത്.

ഗ്രന്ഥസൂചി

[1] ദിവാകരൻ, ആർ.വി.എം. 2014. യന്ത്രവിവർത്തനം. തിരുവനന്തപുരം: കേരളഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട്.

[2] പ്രബോധചന്ദ്രൻ നായർ, വി.ആർ. 2009. ഭാഷാശാസ്ത്രദൃഷ്ടിയിലൂടെ. തിരുവനന്തപുരം: കേരളഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട്.

[3] പ്രബോധചന്ദ്രൻ നായർ, വി.ആർ. 2012. ഭാഷാശാസ്ത്രനിഘണ്ടു. തിരുവനന്തപുരം: കേരളഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട്.

[4] മാധവൻ, പി. 2006. ഭാഷയും സാങ്കേതികവിദ്യയും; സിദ്ധാന്തവും പ്രയോഗവും. വിജ്ഞാനകൈരളി. ഡിസംബർ. വാല്യം-37, ലക്കം-12. തിരുവനന്തപുരം: കേരളഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട്.