മലയാള സർവ്വകലാശാലയ്ക്ക് ഒരു ഭാഷാകമ്പ്യൂട്ടിങ്ങ് രൂപരേഖ

സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ്

ഫെബ്രുവരി 8, 2013 ന് തിരുവനന്തപുരത്ത് മലയാളം യൂണിവേഴ്സിറ്റിയും ഐടിമിഷനും ചേർന്ന് നടത്തുന്ന മലയാളം കമ്പ്യൂട്ടിങ്ങ് ശില്പശാലയിൽ സമർപ്പിക്കുന്നത്

ആമുഖം

ഭാഷയുടെ നിലനിൽപ്പിനും വികാസത്തിനും കമ്പ്യൂട്ടർ സാങ്കേതികവിദ്യയിലെ ഭാഷയുടെ സാന്നിദ്ധ്യം അത്യന്താപേക്ഷിതമാണ്. ഭാഷ നിലനിൽക്കണമെങ്കിൽ അതിന്റെ നിലനിൽപ്പിനുള്ള സങ്കേതങ്ങൾ വികസിക്കണം. ലിപിയും അച്ചടിയും കമ്പ്യൂട്ടിങ്ങും അടക്കമുള്ള സങ്കേതങ്ങളിലൂടെയാണ് ഭാഷ ഇന്നത്തെ കാലഘട്ടത്തിൽ എത്തിയത്. അതിന്റെ നേതൃത്വപരമായ പങ്ക് ഏറ്റെടുക്കുക എന്ന വലിയ ഉത്തരവാദിത്വമാണ് മലയാളസർവ്വകലാശാലയ്ക്ക് മുന്നിലുള്ളത്. മലയാള സർവ്വകലാശാലയുടെ വിശാലലക്ഷ്യങ്ങളിൽ ഭാഷാസാങ്കേതികവിദ്യയുടെ വികാസത്തിനും പ്രയോഗത്തിനും പ്രധാനപ്പെട്ട സ്ഥാനമുണ്ടാവുമെന്ന് പ്രതീക്ഷിക്കുന്നു. ഇത്തരുണത്തിൽ ഇന്ത്യയിലെ ഏറ്റവും വലിയ സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ സന്നദ്ധസംഘമായ സ്വതന്ത്രമലയാളം കമ്പ്യൂട്ടിങ്ങിന്റെ ഭാഗത്തുനിന്ന് ചില അഭിപ്രായനിർദ്ദേശങ്ങൾ മുന്നോട്ടുവയ്ക്കാനാഗ്രഹിക്കുന്നു. സർവ്വകലാശാലയുടെ നിലവിലുള്ള പ്രവർത്തനപദ്ധതിയെ പരിപോഷിപ്പിക്കാനുതകുന്ന മാർഗ്ഗരേഖ എന്ന നിലയിൽ ഇത് ഭാഷാസാങ്കേതികവിദ്യയുടെ വികാസത്തിൽ ജനാധിപത്യപരവും ജനകീയവുമായ ഇടപെടലും മുൻകയ്യും ഉറപ്പുവരുത്തുന്നതിൽ സഹായിക്കുമെന്ന് വിശ്വസിക്കട്ടെ. ഭാഷാകമ്പ്യൂട്ടിങ്ങ് യത്നങ്ങൾക്ക് സമീപഭാവിയിൽ ചെലുത്താനാവുന്ന സ്വാധീനവും അവയുടെ പ്രാധാന്യവും മനസ്സിലാക്കി ദീർഘദൃഷ്ടിയോടു കൂടിയുള്ള ഇടപെടലുകളാണ് ഇന്നിന്റെ ആവശ്യം എന്ന മുഖവുരയോടെ, സൂചകങ്ങളിലേക്ക് കടക്കുകയാണ്.

നിർദ്ദേശങ്ങൾ

1. വരുംകാല മലയാളത്തിന്റെ മാദ്ധ്യമം പ്രധാനമായും ഡിജിറ്റൽ ആണെന്ന് മനസ്സിലാക്കി അതിനുള്ള തയ്യാറെടുപ്പ് ഗൌരവത്തോടെ തുടങ്ങേണ്ടതുണ്ട്. അതിനായി മലയാളഭാഷാ ബോധനത്തിൽ പരിഷ്കാരങ്ങൾ കൊണ്ടുവരേണ്ടതുണ്ട്, കമ്പ്യൂട്ടർ അധിഷ്ഠിത വിദ്യാഭ്യാസത്തിൽ വ്യാകരണം, എഴുത്ത്, വായന, ഭാഷാ സംസ്കരണം (language processing) എന്നിവയുടെ പഠനരീതികൾ ഡിജിറ്റൽ സാങ്കേതിക വിദ്യ ഉൾപ്പെടുത്തി പരിഷ്കരിക്കേണ്ടതുണ്ട്. ഇതിനായി സമാന സർവ്വകലാശാലകളുടെ പാഠ്യപദ്ധതി പരിശോധിക്കുന്നത് നന്നായിരിക്കും.
2. ബിരുദ ബിരുദാനന്തര തലങ്ങളിലെ മലയാളഭാഷാ അദ്ധ്യാപകർക്ക് റിഫ്രഷർ കോഴ്സുകളും പരിശീലനങ്ങളും നടത്തി സമകാലിക ഭാഷാ ഉപയോഗങ്ങളെ (സോഷ്യൽ മീഡിയ, ഡിജിറ്റൽ പ്രസാധനം) പരിചയപ്പെടുത്തുകയും അതുപയോഗിക്കാൻ പ്രേരിപ്പിക്കുകയും ബോധനത്തിന് അവയുടെ സാദ്ധ്യതകളെ ഉപയോഗിക്കാൻ പര്യാപ്തരാക്കുകയും വേണം.
3. കമ്പ്യൂട്ടേഷണൽ ലിംഗ്വിസ്റ്റിക്സ് എന്ന വിഷയത്തിൽ മലയാളത്തിലൂന്നി ഡിഗ്രിതലത്തിലും ബിരുദാനന്തര തലത്തിലും ഉള്ള കോഴ്സുകൾ ഉണ്ടായിരിക്കണം. ഈ വിഷയം മലയാളഭാഷയ്ക്ക് പ്രത്യേക ഊന്നൽ നൽകി വേണം തയ്യാറാക്കാൻ.

4. ഭാഷാ സാങ്കേതികവിദ്യയിൽ ഉണ്ടാകുന്ന പുരോഗതി ജനങ്ങൾക്കിടയിൽ എത്തിക്കാൻ ഒരു വെബ് ജേണൽ തുടങ്ങണം.

5. ഭാഷാസാങ്കേതികവിദ്യ, സ്പീച്ച് ആപ്ലിക്കേഷൻ റിസർച്ച്, വിഷ്വൽ കമ്പ്യൂട്ടിങ്ങ്, സാമീപ്യത (accessibility), തുടങ്ങിയ സുപ്രധാന മേഖലകളിലെല്ലാം തന്നെ പ്രത്യേക വൈദഗ്ദ്ധ്യമുള്ള ഭാഷാ വിദഗ്ദർ, ഭാഷാസാങ്കേതിക വിദഗ്ദർ, എഞ്ചിനീയർമാർ കമ്പ്യൂട്ടർ സയന്റിസ്റ്റുകൾ എന്നിവരെയെല്ലാം ഒത്തൊരുമിപ്പിച്ചുള്ള പ്രത്യേക ഇടപെടലുകൾ ആവശ്യമാണ്. ഈ മേഖലകളിൽ എണ്ണപ്പെട്ട സംഭാവനകൾ നൽകിയ മിക്ക അന്തർദേശീയ തലത്തിലുള്ള സർവ്വകലാശാലകളിലും ഇവയ്ക്കായി പ്രത്യേകം പ്രത്യേകം സെന്ററുകളും അവയുടെ ഒത്തുചേർന്നുള്ള ഇടപെടലും കാണാവുന്നതാണ്

6. ഇതോടൊപ്പം മറ്റ് അക്കാദമിക് സ്ഥാപനങ്ങളിലെ ഭാഷാവകുപ്പുകളും എഞ്ചിനീയറിങ്ങ് കോളേജുകളും ഉൾപ്പെടെയുള്ള സ്ഥാപങ്ങളിലെ കർമ്മശേഷി ഗവേഷണത്തിന് ഉപയോഗയോഗ്യമാകുന്ന രീതിയിൽ കൊണ്ടുവരുന്നതിനും ഭാഷാ സർവ്വകലാശാലയ്ക്കു് സാധിക്കേണ്ടതാണ്

7. മറ്റു് സർവ്വകലാശാലകളിലും സ്ഥാപനങ്ങളിലും ഉള്ള ഗവേഷകർക്ക് , ഭാഷാസാങ്കേതികവിദ്യയുമായി ബന്ധപ്പെട്ട വിഷയങ്ങളിൽ സാങ്കേതികവിദ്യയുടെ വികസനത്തിന് ഉപയോഗപ്രദമായ രീതിയിലുള്ള ഗവേഷണ വിഷയങ്ങളിൽ ശ്രദ്ധകേന്ദ്രീകരിക്കാനും അതിനായി മലയാളം യൂണിവേഴ്സിറ്റിയിലെ സൌകര്യങ്ങളുപയോഗിക്കാനും വേണ്ടി ഫെല്ലോഷിപ്പ് സമ്പ്രദായം ഏർപ്പെടുത്താവുന്നതാണ്

8. കഴിഞ്ഞ 10 വർഷമായി മലയാളം ഭാഷാകമ്പ്യൂട്ടിങ്ങിൽ നടന്ന മിക്ക സാങ്കേതിക മുന്നേറ്റങ്ങളും വ്യക്തികളുടേയോ പൊതുസമൂഹത്തിലെ സ്വതന്ത്രമലയാളം കമ്പ്യൂട്ടിങ്ങ് അടക്കമുള്ള അറിവുക്കൂട്ടങ്ങളുടെയോ (knowledge societies) മുൻകൈയിൽ ആയിരുന്നു. ഇത്തരം community expertise ധാരാളമുള്ള ഒരു മേഖലയാണ് ഭാഷാസാങ്കേതിക വിദ്യ. ഈ രംഗത്തെ വൈദഗ്ദ്ധ്യങ്ങളെയും സംഭാവനകളേയും ഉപയോഗപ്പെടുത്താനും പ്രോത്സാഹിപ്പിക്കാനും പരിപോഷിപ്പിക്കാനും ഉള്ള പരിശ്രമങ്ങൾ മലയാളം സർവ്വകലാശാലയുടെ ഭാഗത്തുനിന്ന് ഉണ്ടാവണം.

9. ഭാഷാ കമ്പ്യൂട്ടിങ്ങ് രംഗത്തു് വ്യക്തികളും പ്രസ്ഥാനങ്ങളും സ്ഥാപനങ്ങളും റിസർച്ച് ഇൻസ്റ്റിറ്റ്യൂഷനുകളും ഒത്തുചേർന്ന് നടത്തുന്ന interdisciplinary പ്രവർത്തനങ്ങൾ വളരെ സുപ്രധാനമാണ്. അതിനുള്ള സാഹചര്യമൊരുക്കണം

10. സ്പീച്ച് പ്രൊസസിങ്ങ്, ഇമേജ്/കാരക്റ്റർ റെക്കഗ്നിഷൻ, മെഷീൻ ട്രാൻസിലേഷൻ എന്നീ രംഗങ്ങളിൽ പ്രവർത്തിക്കുന്ന ഗവേഷകർക്കും സ്വതന്ത്രസോഫ്റ്റ്വെയർ സമൂഹങ്ങൾക്കും പ്രോഗ്രാമർമാർക്കും സ്വതന്ത്രമായ ഡാറ്റാ കോർപ്പസ്സുകൾ വളരെ അത്യാവശ്യമായ ഒന്നാണ്. സ്പീച്ച് ഇമേജ് കോർപ്പസ്സുകളുടെ ലഭ്യത സ്വതന്ത്ര ലൈസൻസിൽ ഉറപ്പുവരുത്തുകയാണെങ്കിൽ ഈ രംഗത്തു് വലിയ മുന്നേറ്റങ്ങൾ നടക്കും. ഇന്ന് മലയാളത്തിൽ ഒരൊറ്റ സ്വതന്ത്ര കോർപ്പസ്സുപോലും ലഭ്യമല്ലാത്ത അവസ്ഥയുണ്ട്. മലയാളം സർവ്വകലാശാലക്ക് ഈ പ്രശ്നം പരിഹരിക്കാനാവണം.

11. നിരവധി ഭാഷാഭേദങ്ങളും ഉപഭാഷകളും (ആദിവാസി മലയാളങ്ങൾ , അറബി മലയാളം തുടങ്ങിയവ) നിലനിൽക്കുന്ന സ്ഥലമാണ് കേരളം. ഇത്തരം ഭാഷാഭേദങ്ങളുടെയും ഉപഭാഷകളുടെയും കമ്പ്യൂട്ടിങ്ങ് സാധ്യതകളിലുള്ള ഗവേഷണങ്ങൾ നടക്കേണ്ടതുണ്ട്

12. ഇന്ത്യൻ ഭാഷകളുടെ കമ്പ്യൂട്ടിങ്ങിൽ ഇന്നു പ്രവർത്തിക്കുന്നതു് നിരവധി സ്ഥാപനങ്ങളാണ്. അക്കാദമിക് സ്ഥാപനങ്ങളും സ്വതന്ത്രസോഫ്റ്റ്‌വെയർ സമൂഹങ്ങളും, കോർപ്പറേറ്റ് കമ്പനികളും, അർദ്ധസർക്കാർ സ്ഥാപനങ്ങളും പ്രസാധകരും നിരവധി കൊച്ചു സംരംഭകരും പ്രവർത്തിക്കുന്ന രംഗമാണ് ഇതു്. ഈ രംഗത്തെ ഗവേഷണങ്ങളും നടന്നുവരുന്ന പ്രവർത്തനങ്ങളും മനസ്സിലാക്കേണ്ടതു് സുപ്രധാനമായതിനാൽ ഈ രംഗത്തുള്ള ഏവരേയും പങ്കെടുപ്പിച്ചുകൊണ്ട് കൊല്ലത്തിലൊരിക്കൽ ഒരു ഇൻഡിക് കമ്പ്യൂട്ടിങ്ങ് കോൺഫറൻസ് നടത്തേണ്ടതാണ്. പുതിയ സഹവർത്തിത്വത്തിന്റെ സാധ്യതകൾ തുറക്കാനും ഇതു വഴിയൊരുക്കും.

13. പല സർക്കാർ ഗവേഷണ സ്ഥാപനങ്ങളിലും നിർമ്മിക്കപ്പെടുന്ന സമൂഹത്തിന് ഉപകാരപ്രദമാവേണ്ട പല സോഫ്റ്റ്‌വെയറും സ്വതന്ത്ര ലൈസൻസിങ്ങ് നയം സ്വീകരിയ്ക്കാത്തതിനാൽ ജനങ്ങൾക്ക് ഉപയോഗയോഗ്യമല്ലാതെ കിടക്കുന്ന അവസ്ഥ ഇന്നുണ്ട്. മലയാളം സർവ്വകലാശാല നേരിട്ടോ /മറ്റു് ഏജൻസികൾ മുഖേനയോ നിർമ്മിക്കുന്ന സോഫ്റ്റ്‌വെയർ സ്വതന്ത്രസഫ്റ്റ്വെയറുകളായിരിക്കുമെന്ന്

നയപരമായ തീരുമാനം എടുക്കണം. സോഫ്റ്റ്‌വെയർ നിർമ്മാണത്തിന് ഗിറ്റ് അധിഷ്ഠിത പബ്ലിക് റെപ്പോസിറ്ററികൾ ഉപയോഗപ്പെടുത്തുന്ന മാതൃക ഇന്ന് പല വിദേശ സർവ്വകലാശാലകളും പിന്തുടരുന്നുണ്ട്

14. ടൈപ്പ് ഫേസ് ഡിസൈനിങ്ങ് പ്രത്യേകമായിത്തന്നെ മലയാളം സർവ്വകലാശാല പരിഗണിക്കേണ്ട ഒരു വിഷയമാണ്. ലാറ്റിനേതര ഭാഷകളുടെ ടൈപ്പ് ഫേസ് ഡിസൈനിങ്ങിലെ വിദഗ്ധരുടെ ക്ഷാമം പരിഹരിക്കാൻ ഇത്തരം ഇടപെടലുകളിലൂടെ കഴിയും. മലയാളത്തിലെ ലിപി രൂപങ്ങളുടെ വൈവിധ്യമാർന്ന നിരവധി എഴുത്തുരീതികൾ (ചുവരെഴുത്തു രീതികൾ, ബാനർ എഴുത്തുരീതികൾ തുടങ്ങിയവ) അപ്രത്യക്ഷമായിപോകാനനുവദിയ്ക്കാതെ ഡിജിറ്റൽ രൂപത്തിലാക്കി സംരക്ഷിക്കലും ഈ രംഗത്തെ പ്രവർത്തനങ്ങളുടെ ഭാഗമാക്കണം.
15. മലയാളത്തിൽ വിവരവ്യവസ്ഥാ നിർമ്മിതി നേരിടുന്ന പ്രശ്നങ്ങളെക്കുറിച്ച് (Information extraction, retrieval and presentation) പഠിക്കാനും പരിഹാരങ്ങളുണ്ടാക്കാനും ഉള്ള ശ്രമങ്ങൾ നടക്കേണ്ടതുണ്ട്
16. മലയാളം സർവ്വകലാശാലയിൽ തയ്യാറാക്കപ്പെടുന്ന റിസർച്ച് പേപ്പറുകളുടെ ഒരു ഓപ്പൺ ആക്സസ് റെപ്പോസിറ്ററി ഉണ്ടാവേണ്ടതാണ്. ഉദാ: http://eprints.iisc.ernet.in/
17. മലയാളം സർവ്വകലാശാലയിലെ ഗവേഷണ, പഠന ഫലങ്ങൾ സമൂഹത്തിൽ ഫലപ്രദമായും വേഗത്തിലും എത്തിക്കുവാനുമുള്ള സംവിധാനങ്ങളൊരുക്കണം.

മലയാള ഭാഷാ കമ്പ്യൂട്ടിങ്ങിന് വേണ്ട സെന്ററുകൾ

1 Centre for Language Processing

ഈ സെന്റർ താഴെപ്പറയുന്ന വിഷയങ്ങളിൽ സജീവമായി പ്രവർത്തിക്കണം

- Natural language processing and computational linguistics,
- Information extraction, retrieval and presentation in Indian languages
- Big data analytics
- machine learning
- machine translation
- text analysis.

2. Centre for Visual Information Technology

Basic and advanced research in image processing, computer vision and computer graphics.
ഈ സെന്റർ താഴെപ്പറയുന്ന വിഷയങ്ങളിൽ സജീവമായി പ്രവർത്തിക്കണം

- Document image processing especially searching in digitized books in Indian languages
- Researches on Image recognition on palm prints, handwriting, and hand geometry
- Content-based annotation and search of large collections of images and videos

3. Centre for Speech Processing

ഈ സെന്റർ താഴെപ്പറയുന്ന വിഷയങ്ങളിൽ മലയാളത്തിൽ പ്രത്യേക ഊന്നലോടെ സജീവമായി പ്രവർത്തിക്കണം

- acoustic processing
- automatic speech recognition
- speech translation
- phonetic engine for Indian languages
- speaker recognition
- dialog systems in speech mode
- speech synthesis
- speech enhancement
- speech coding
- information retrieval in speech mode,

4. Center for Educational and assistive technology

ഈ സെന്റർ താഴെപ്പറയുന്ന വിഷയങ്ങളിൽ മലയാളത്തിൽ പ്രത്യേക ഊന്നലോടെ സജീവമായി പ്രവർത്തിക്കണം

- Assistive web technologies.
- Localisation and Internationalisation Features
- Interdisciplinary on speech and visual systems on accessibility angle.
- Technology devices and software for assisting diffentially abled persons.
- Educational systems for Learning disability and language learning

5. Centre for Typeface design

- Indian Language font design
- മലയാള ലിപി വൈവിധ്യ പഠനവും അവയുടെ ഫോണ്ട് രൂപത്തിലുള്ള സംരക്ഷണവും
- ലിപി പരിണാമ പഠനം

കുറിപ്പ്: ഈ രേഖ ഭാഷാകമ്പ്യൂട്ടിങ്ങ് രംഗത്ത് മലയാളം സർവ്വകലാശാല നടത്തേണ്ട പ്രാഥമികമായ ഇടപെടലുകളുടെ രൂപരേഖമാത്രമാണ്, ഒരു സമ്പൂർണ്ണ ആക്ഷൻ പ്ലാനല്ല.