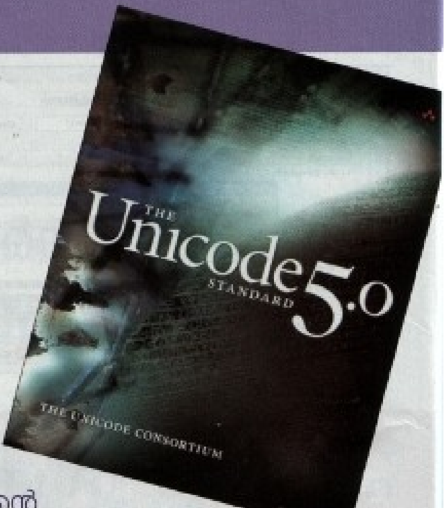
സന്തോഷ് തോട്ടിങ്ങൽ

യൂണികോഡ്: മലയാളത്തിന് സംഭവിക്കുന്നതെന്ത്?



ഭാഷയുടെ അടിസ്ഥാന തത്വങ്ങളെ വിസ്മരിച്ച് മലയാളത്തിന്റെ ഡിജിറ്റൽ ഭാവിയെത്തന്നെ അപകടത്തിലാക്കുന്ന തീരുമാനങ്ങളാണ് യൂണികോഡ് കൺസോർഷ്യം പലപ്പോഴും കൈക്കൊള്ളുന്നതെന്ന് ലേഖകൻ ചൂണ്ടിക്കാട്ടുന്നു.

യൂണികോഡ് എന്നത് ലോകത്താകമാനമുള്ള ഭാഷകളെ ഡിജിറ്റൽ രൂപത്തിൽ സംരക്ഷിക്കാനും സംസ്കരിക്കാനുമുള്ള ഒരു ഏകീകൃത സ്റ്റാൻഡേർഡ് ആണ്. എല്ലാ ഭാഷകളിലെയും അക്ഷരങ്ങൾക്ക് അതിനു മാത്രം സ്വന്തമായ ഒരു കോഡ് പോയിന്റ് നിർവചിക്കുകയും, അതുവഴി വിവരങ്ങളുടെ കൈമാറ്റം സുഗമമാക്കുകയും ചെയ്യുന്ന ഒരു സംവിധാനമാണത്. തൊണ്ണൂറിലധികം ലിപികളിലായി ഒരു ലക്ഷത്തിലധികം ലിപിരൂപങ്ങളെ ഇങ്ങനെ യൂണികോഡ് നിർവചിച്ചിരിക്കുന്നു. പല സ്ഥലങ്ങളിലും നിലനില്ക്കുന്ന വ്യത്യസ്ഥ ഡാറ്റാ എൻകോഡിങ് മാനകങ്ങളെ ഒറ്റ മാനകമായ യൂണികോഡിലേക്ക് കൊണ്ടുവരാൻ യൂണികോഡ് കൺസോർഷ്യം ലക്ഷ്യമിടുന്നു. ഇന്റർനെറ്റിലെയും കംപ്യൂട്ടറുകളിലെയും മലയാളഭാഷ യൂണികോഡ് മാനകമനുസരിച്ചുള്ള എൻകോഡിങ് രീതിയാണ് പിന്തുടരുന്നത്.

യൂണികോഡ് കൺസോർഷ്യം എന്ന ഓർഗനൈസേഷൻ പ്രമുഖ കംപ്യൂട്ടർ/സോഫ്റ്റ്‌വെയർ ഉത്പാദകക്കമ്പനികൾ രൂപം കൊടുത്തതും നോക്കി നടത്തുന്നതുമായ സ്ഥാപനമാണ്. ഭാഷയെ സംബന്ധിച്ചിട്ടുള്ളതാണെങ്കിലും, ഭാഷാശാസ്ത്രപരമായ നിയമങ്ങളോ, അക്കാദമിക് നിരീക്ഷണങ്ങളോ യൂണികോഡിന്റെ ഭാഷാപരമായ തീരുമാനങ്ങൾക്ക് അടിസ്ഥാനമാവാറില്ല. അതിലെ അംഗങ്ങളുടെ സോഫ്റ്റ്‌വെയർ ഉത്പന്നങ്ങൾക്കു വേണ്ട കാര്യങ്ങൾക്കാണ് പ്രാമുഖ്യം കൊടുക്കാറുള്ളത്. ഭാഷയിലെ നിലനിൽക്കുന്ന ലിപിരൂപങ്ങളെ എൻകോഡ് ചെയ്യുക മാത്രമാണ് പ്രഖ്യാപിത ലക്ഷ്യമെങ്കിലും അവയെങ്ങനെ എഴുതണമെന്ന ഭാഷാശാസ്ത്രപരമായ അടിസ്ഥാനതത്വങ്ങളെ വിസ്മരിക്കുന്നത് ഭാഷയുടെ ഡിജിറ്റൽ നിലനിൽപ്പ് അപകടത്തിലാക്കുന്നു. ഇതുകൂടാതെ, ഭാഷയെ സംബന്ധിച്ച സുപ്രധാന തീരുമാനങ്ങളെടുക്കുന്നത് ഒട്ടും തന്നെ സുതാര്യമല്ലാത്ത നടപടികളിലൂടെയാണു താനും. മലയാളത്തിന്റെ ഭാവിയിലെ നിലനിൽപ്പ് ഡിജിറ്റൽ രൂപത്തിലാണ്. ആ മലയാളം എങ്ങനെ എഴുതണമെന്ന് മലയാളം മാതൃഭാഷ ആയവരോ, അതിൽ പ്രാവീണ്യമുള്ളവരോ ആയ മലയാളികൾക്കും വിദഗ്ധർക്കും ഒട്ടും പ്രാതിനിധ്യമില്ലാതെ സോഫ്റ്റ്‌വെയർ കമ്പനികളുടെ പ്രതിനിധികൾ അവർക്കു വേണ്ട വിധത്തിൽ നിർവചിക്കാൻ തുടങ്ങുകയും അതു സ്റ്റാൻഡേഡ് ആവുകയും ചെയ്യുമ്പോൾ നിരവധി പ്രശ്നങ്ങൾ ഉണ്ടാവുന്നു. ഈ പ്രശ്നങ്ങൾ മലയാളികളുടെ ശ്രദ്ധയിൽ കൊണ്ടുവരാനും ചുരുങ്ങിയത് ഇങ്ങനെയൊക്കെ സംഭവിക്കുന്നുണ്ടെന്ന് അറിഞ്ഞിരിക്കുന്നതിനും വേണ്ടിയാണ് ഈ ലേഖനം.

മലയാളത്തെ സംബന്ധിച്ചിടത്തോളം സുപ്രധാനമായ ഒരു തീരുമാനം യൂണികോഡ് എടുത്തതിന്റെ ഉള്ളറകളാണ് ഇനി പറയാൻ പോകുന്നത്. ആദ്യം സ്വല്പം വ്യാകരണം. ഏവർക്കുമറിയാവുന്ന മലയാളത്തിലെ "ന്റ" എന്ന കൂട്ടക്ഷരം ന + ് + റ എന്ന രീതിയിലാണ് നാളിതുവരെ യൂണികോഡിൽ എഴുതിപ്പോന്നിരുന്നത്. ഇതിന്റെ അടിസ്ഥാനമെന്താണെന്നു ചോദിച്ചാൽ അതിങ്ങനെയാണ്.

ക, ച, ട, ത, പ എന്നീ വ്യഞ്ജന വർഗ്ഗങ്ങളിൽ ഓരോ വർഗ്ഗത്തിലെയും ഖരാക്ഷരവും അനുനാസികവും ചേർന്ന് കൂട്ടക്ഷരമുണ്ടാവും. ഉദാഹരണം:

ങ + ് + ക = ങ്ക
ഞ + ് + ച = ഞ്ച
ണ + ് + ട = ണ്ട
ന + ് + ത = ന്ത
മ + ് + പ = മ്പ എന്നിങ്ങനെ

ക, ച, ട, ത, പ എന്നീ അഞ്ചു വർഗ്ഗങ്ങൾ കൂടാതെ മലയാളത്തിൽ ഉണ്ടായിരുന്നെന്ന് പറയപ്പെടുന്ന ഒരു വർഗ്ഗമാണ് റ്റ വർഗ്ഗം. ശരിക്കും റ്റ അല്ല അതിന്റെ പകുതി അക്ഷരത്തിന്റെ വർഗ്ഗമാണത്. റ്റ എന്നത് ഇപ്പോളില്ലാത്ത ആ അക്ഷരത്തിന്റെ കൂട്ടക്ഷരമാണ് (ഇരട്ടിപ്പ്). ഇംഗ്ലീഷിലെ hat, cat എന്നൊക്കെ പറയുമ്പോഴുള്ള ആ ശബ്ദം തന്നെ. ട യുടെയും ത യുടെയും ഇടയിലാണ് ഇതിന്റെ ഉച്ചാരണം വരുന്നത് (ഉച്ചരിച്ചു നോക്കുക-ക മുതൽ പ വരെ പറയുമ്പോൾ നാവിന്റെ ചലനം ശ്രദ്ധിക്കുക). ലിപിദാരിദ്ര്യം എന്നും മറ്റും കേരളപാണിനീയത്തിൽ പറയുമ്പോലെ, ഈ അക്ഷരമെഴുതാൻ മലയാളത്തിൽ ലിപിയില്ല. പക്ഷേ ഇതിന്റെ അനുനാസികം നാം വളരെയേറെ ഉപയോഗിക്കുന്ന ഒന്നാണ്.

Table I Unique numbers for scripts assigned by The Unicode Consortium

Script	Assigned unique number
Arabic	U+0600 – U+06FF (01536 – 01791)
Devnagari	U+0900 – U+097F (02304 – 02431)
Bengali	U+0980 – U+09FF (02432 – 02559)
Gurumukhi	U+0A00 – U+0A7F (02560 – 02687)
Gujarati	U+0A80 – U+0AFF (02688 – 02815)
Oriya	U+0B00 – U+0B7F (02816 – 02943)
Tamil	U+0B80 – U+0BFF (02944 – 03071)
Telugu	U+0C00 – U+0C7F (03072 – 03199)
Kannada	U+0C80 – U+0CFF (03200 – 03327)
Malayalam	U+0D00 – U+0D7F (03328 – 03455)

അതാണ് പനയിലെ ന, നരിയിലെ ന അല്ല. മലയാളത്തിനില്ലെങ്കിലും തമിഴിൽ ഇതിന് ലിപിയുണ്ട്. "ன"

അനുനാസികമായ പനയിലെ നയും ഖരാക്ഷരവും ചേർന്നതാണ് ന്റ. പനയിലെ നയ്ക്ക് ന തന്നെ എടുക്കുകയും, റ്റയുടെ പകുതി ഖരത്തിനെടുക്കുകയും ചെയ്തപ്പോഴാണ് ന്റ വന്നത് (ഈ വിവരങ്ങൾക്ക് കടപ്പാട്: കേരളപാണിനീയം)

മൈക്രോസോഫ്റ്റിന്റെ മലയാളം ഫോണ്ടാണ് കാർത്തിക. മലയാളത്തിലെ മോശപ്പെട്ട ഫോണ്ടുകളിലൊന്നാണ് ഈ ടൈപ്പ് റൈറ്റർ ലിപി കുത്തക ഫോണ്ട്. വിൻഡോസിൽ സ്വതവേ മലയാളം കാണിക്കുന്നതും ഇതുവെച്ചാണ്. ഈ ഫോണ്ടിൽ ന്റ എന്നെഴുതിയാൽ ൻറ എന്നേ കാണുള്ളൂ (കാർത്തിക ഫോണ്ടുപയോഗിക്കുന്നവർക്കിതു മനസ്സിലാവും.) ആ ഫോണ്ടിന്റെ പിഴവായിരുന്നു അത്. ന്റ എന്നതിന്റെ ചർച്ച യൂണിക്കോഡിൽ വന്നപ്പോൾ മൈക്രോസോഫ്റ്റ് പ്രതിനിധികൾ അവരുടെ ഫോണ്ട് നിലവിൽ ന്റ, ൻറ എന്നു കാണിക്കുന്നതുകൊണ്ട് ൻ+റ എന്ന നിയമം പറ്റില്ലെന്നും പകരം ൻ+ ്+റ എന്നുവേണമെന്നും ആവശ്യപ്പെട്ടു. യൂണികോഡ് അതങ്ങ് അംഗീകരിക്കുകയും ചെയ്തുവെന്നാണ് അറിയാൻ കഴിഞ്ഞത്. യൂണികോഡിന്റെ 5.1 പതിപ്പിൽ ഈ രീതിയാണ് ന്റ എന്നതിന് നിർവചിച്ചിരിക്കുന്നത്.

ചുരുക്കിപ്പറഞ്ഞാൽ തങ്ങളുടെ ഒരു പിഴവ് തിരുത്തുന്നതിനുപകരം മലയാളഭാഷയെ മൈക്രോസോഫ്റ്റ് മാറ്റിമറിച്ചു എന്നർഥം. ഈ സ്റ്റാൻഡേഡ് ആണ് ഇനി സോഫ്റ്റ്‌വെയറുകൾ പിന്തുടരേണ്ടത്. പുതിയ ൻ+ ്+റ എന്നതിലെ പിഴവെന്താണെന്നു ചോദിച്ചാൽ ചില്ലക്ഷരങ്ങൾ ചന്ദ്രക്കലയുടെ സഹായത്തോടെ കൂട്ടക്ഷരമുണ്ടാക്കില്ലെന്ന ഭാഷാ നിയമം തെറ്റിച്ചു. പോരാത്തതിന് ഇതുവരെ മലയാളികളെഴുതിയ ന്റ കളെല്ലാം ഈ സ്റ്റാൻഡേർഡോടെ അസാധുവായി.

അക്ഷരങ്ങളുടെ എൻകോഡിങ്ങല്ലാതെ കൂട്ടക്ഷരങ്ങളുടെ എൻകോഡിങ്ങിൽ യൂണിക്കോഡ് കൈകടത്തുന്നതും വളരെ അപൂർവതയാണ്.

നേരത്തെ പറഞ്ഞ റ്റയുടെ ഒറ്റയ്ക്കും. പനയിലെ നയ്ക്കും മലയാളത്തിൽ ലിപിയില്ലെന്ന് എ.ആർ.രാജരാജവർമ കേരളപാണിനീയത്തിൽ പറയുന്നു. പക്ഷേ ഈ ലിപികൾ മലയാളത്തിലില്ലെന്നും ഇങ്ങനെ ഒരു വർഗം തമിഴിലും മലയാളത്തിന്റെയും തമിഴിന്റെയും മൂലലിപിയായ ഗ്രന്ഥലിപിയിലും ഉണ്ടെന്ന് പറയാനായി രാജരാജവർമ കേരളപാണിനീയത്തിൽ രണ്ട് അക്ഷരങ്ങൾ ഉപയോഗിച്ചിരുന്നു. ഗ്രന്ഥലിപിയിൽ നിന്നും കടമെടുത്താണ് രാജരാജവർമ ആ അക്ഷരങ്ങളെ പരിചയപ്പെടുത്തിയത് (ചിത്രം കാണുക). രാജരാജവർമ ഇത് വിശദീകരിക്കാനായി എടുത്ത ആ രണ്ട് അക്ഷരങ്ങൾ മലയാളത്തിലുള്ളതാണെന്നും അവ മലയാള അക്ഷരങ്ങളായി യൂണികോഡ് എൻകോഡിങ്ങിൽ ചേർക്കണം എന്നും മൈക്കൽ എവേഴ്സൺ എന്ന ലിപി ഗവേഷകൻ യൂണികോഡ് കൺസോർഷ്യത്തോടാവശ്യപ്പെടുകയും യൂണികോഡ് അതംഗീകരിക്കുകയും ചെയ്തിരിക്കുന്നു. യൂണികോഡിന്റെ അടുത്ത പതിപ്പിൽ ഇവ ഔദ്യോഗികമായി പുറത്തുവരും. പക്ഷേ യൂണികോഡ് അംഗീകാരം കിട്ടിയ ശേഷമാണ് മലയാളികൾ ഇതേപ്പറ്റി അറിയുന്നത്. ഇപ്പോഴാണ് ഇതിനെപ്പറ്റിയുള്ള ചർച്ചകൾ നടക്കുന്നത്. അംഗീകാരം വന്ന ശേഷം ഈ ലിപികൾ ഉൾപ്പെടുത്താൻ നിർദേശിച്ചവർ തങ്ങൾ റ്റ-യുടെ പകുതിക്കു നിർദേശിച്ച അക്ഷരം തമിഴിലെ 'ടി' ആയിരുന്നെന്നും തെറ്റിയ അക്ഷരം തിരുത്തണമെന്നും ഇപ്പോൾ ആവശ്യപ്പെടുന്നു. ഇതിൽ നിന്നും എത്ര നിരുത്തരവാദപരമായാണ് ഭാഷയുടെ ഭാവിയെപ്പറ്റിയുള്ള തീരുമാനങ്ങൾ എടുക്കപ്പെടുന്നത് എന്ന് വ്യക്തമാണല്ലോ. പി.വിശാലാക്ഷി എഴുതിയ ഗ്രന്ഥലിപിയെപറ്റിയുള്ള പുസ്തകമാണ് ഇപ്പോൾ മലയാളത്തിൽ ഈ അക്ഷരങ്ങൾ ഉണ്ടായിരുന്നു എന്നതിന് നിദർശനമായി എടുത്തുകാണിക്കപ്പെടുന്നത്. ഗ്രന്ഥലിപി

37. ഺ, ഩ എന്ന ഺ വർഗ്ഗമാകട്ടെ, കവർഗ്ഗചവർഗ്ഗാദികൾപോലെ സ്പർശങ്ങളിലുള്ള വർഗ്ഗംതന്നെ. ഺ ഖരം; ഩ അനുനാസികം. അതിമൃദുഘോഷങ്ങൾ തമിഴിൽ ഇല്ലല്ലോ. അതിനാൽ ഖരവും അനുനാസികവും മാത്രമേ ഉള്ളു. ഈ വർഗ്ഗത്തിന് സ്ഥാനം ദന്തമൂലത്തിനും മൂർദ്ധാവിനും മദ്ധ്യേ ഉള്ള (ഊനിന് അടുത്ത) പിൻഭാഗം എന്ന പ്രദേശം; അതിനാൽ ഈ വർഗ്ഗം വർത്സ്യം. അക്ഷരമാലയുടെ രചന ഉള്ളിൽനിന്നും വെളിയിലേക്കുള്ള സ്ഥാനക്രമംപ്രമാണിച്ചു ചെയ്തിട്ടുള്ളതാകയാൽ വർത്സ്യമായ ഺ വർഗ്ഗത്തെ മൂർദ്ധന്യമായ ടവർഗ്ഗം കഴിഞ്ഞ് ദന്ത്യമായ തവർഗ്ഗത്തിനു മുൻപായി പഠിക്കേണ്ടതായിരുന്നു; തമിഴുവൈയാകരണന്മാരാകട്ടെ, സംസ്കൃതാക്ഷരങ്ങളെ എല്ലാം മുറയ്ക്ക് എടുത്തതിന്റെശേഷം ഺ, ഩ എന്ന ഈ വർഗ്ഗത്തെ ഉള്ളതിലും ഒടുവിൽ തള്ളിക്കളഞ്ഞുവെന്നേ ഉള്ളൂ.

ഺകാരത്തിന്

എന്താണ് യൂണികോഡ്?

കമ്പ്യൂട്ടറിനുള്ളിൽ എല്ലാം സംഖ്യകളാണ്. പൂജ്യത്തിന്റെയും ഒന്നിന്റെയും കൂട്ടമായാണ് കമ്പ്യൂട്ടർ വിവരങ്ങൾ മനസ്സിലാക്കുന്നത്. കമ്പ്യൂട്ടറിന് അക്ഷരങ്ങളും ചിഹ്നങ്ങളും സ്വീകരിക്കണമെന്നുണ്ടെങ്കിൽ അവയെ സംഖ്യാരൂപത്തിലാക്കണം. അതിനുള്ള മാർഗമാണ് 'എൻകോഡിങ്'. ഓരോ അക്ഷരത്തിന്റെയും സംഖ്യാരൂപത്തിന് 'ക്യാരക്ടർ കോഡിങ്' എന്നു പറയും. ഇത്തരത്തിലുള്ള ഒരു കോഡാണ് ആസ്കീ (ASCII - American Standard Code for Information Interchange) ഇത് ഉപയോഗിച്ച് ഉണ്ടാക്കാവുന്ന അക്ഷരാദികളുടെ എണ്ണം പരമാവധി 256 ആണ്. ഇതിൽ ആദ്യത്തെ 128 എണ്ണം ഇംഗ്ലീഷിനു നൽകിയിരിക്കുന്നു. ഉദാഹരണത്തിന്, ഇംഗ്ലീഷിൽ A ascii code 1000001 എന്നും a-യുടെത് 1100001 എന്നുമാണ്. എന്നാൽ ആസ്കീ കോഡിൽ ലോകത്തെ പ്രധാന ഭാഷകളെയെല്ലാം ഉൾപ്പെടുത്താൻ കഴിയില്ല. അതിന് പുതിയ കോഡിങ് സിസ്റ്റം ആവശ്യമായി വന്നു. എല്ലാം ഭാഷകളിലെയും ഓരോ അക്ഷരവും ചിഹ്നവും മറ്റൊന്നിനും ഇല്ലാത്ത തനതായ ഒരു ബൈനറി സംഖ്യയായി നിജപ്പെടുത്തിയാൽ അത് പിൻതുടരുന്ന ഏത് കമ്പ്യൂട്ടറിനും എല്ലാ ഭാഷകളെയും ഒന്നിച്ച് മനസ്സിലാക്കാൻ കഴിയും. ഇതാണ് യൂണികോഡ്. മലയാളത്തിന്റെ കാര്യം തന്നെ നോക്കാം. ഇതുവരെയുള്ള ഐ.എസ്.എം പോലുള്ള സോഫ്റ്റ്‌വെയറുകളുപയോഗിച്ച് മലയാളം അക്ഷരം ടൈപ്പ് ചെയ്യാൻ കഴിയുമെങ്കിലും ഓരോ മലയാള അക്ഷരത്തെയും കമ്പ്യൂട്ടർ അതുമായി ബന്ധിച്ചുള്ള ഒരു ആസ്കി അഥവാ ഇംഗ്ലീഷ് അക്ഷരമോ ചിഹ്നമോ ആയിട്ടാണ് മനസ്സിലാക്കിയിരുന്നത്. അതുകൊണ്ടു തന്നെ ആ പ്രത്യേക സോഫ്റ്റ്‌വെയർ ഇൻസ്റ്റാൾ ചെയ്യാത്ത കമ്പ്യൂട്ടറിൽ മലയാള അക്ഷരങ്ങൾക്ക് പകരം അതിനോട് ബന്ധപ്പെട്ട ഇംഗ്ലീഷ് അക്ഷരങ്ങളാവും പ്രത്യക്ഷപ്പെടുക. എന്നാൽ യൂണികോഡ് പ്രകാരം ഏത് ഭാഷയിലെ അക്ഷരങ്ങളെയും കമ്പ്യൂട്ടർ അതിനുമാത്രമുള്ള ബൈനറി സംഖ്യകളായി മനസ്സിലാക്കുന്നതിനാൽ ആ ഭാഷയിലെ അക്ഷരങ്ങളായി തന്നെ സ്ക്രീനിൽ പ്രത്യക്ഷപ്പെടും.

യൂണിക്കോഡ് വേറെ എൻകോഡ് ചെയ്യാൻ നടപടികൾ എടുക്കുന്നുണ്ട്.

മലയാളത്തിലെ ചില്ലക്ഷരങ്ങളെപ്പറ്റിയുള്ള ദീർഘകാലത്തെ വിവാദത്തിനൊടുവിൽ ചില്ലക്ഷരങ്ങളെ വേറിട്ട അക്ഷരങ്ങളായി പരിഗണിച്ച് യൂണിക്കോഡ് സ്റ്റാൻഡേഡ് 5.1 ൽ പരിഷ്കരിച്ചു. കേരള യൂണിവേഴ്സിറ്റിയിൽ നടന്ന ചർച്ചയിൽ ഭാഷാവിദഗ്ധരുടെയും സാങ്കേതികപ്രവർത്തകരുടെയും ഈ പ്രശ്നത്തെ സംബന്ധിച്ച നിർദ്ദേശങ്ങൾ യൂണിക്കോഡ് തള്ളുകയായിരുന്നു. അതോടുകൂടി ഇതുവരെ നിലവിലുണ്ടായിരുന്നതും ഇപ്പോൾ പലരും എഴുതിക്കൊണ്ടിരിക്കുന്നതുമായ രീതിയിലുള്ള ചില്ലക്ഷരങ്ങളും പുതിയ ചില്ലക്ഷരങ്ങളും ചേർന്ന് മലയാളം രണ്ടുവിധമായി. പഴയതും പുതിയതുമായ ചില്ലുകൾ തുല്യമല്ലാത്തതിനാൽ തെരച്ചിൽ, താരതമ്യം, സോർട്ടിങ്ങ് തുടങ്ങിയവയിലെല്ലാം പ്രശ്നങ്ങളായിരിക്കുന്നു. പതിനൊന്നായിരം ലേഖനങ്ങളുള്ള മലയാളം വിക്കിപീഡിയ ഇതിലേതുരീതി പിന്തുടരണമെന്നും അങ്ങനെ ഒരു രീതി പിന്തുടർന്നാൽ മറ്റേ രീതിയിലുള്ള വിവരങ്ങളെ എങ്ങനെ പിന്തുണയ്ക്കുമെന്നും ഉള്ള പ്രശ്നത്തിൽ കുടുങ്ങിക്കിടക്കുന്നു.

ഇപ്പോൾ യൂണിക്കോഡിന്റെ പരിഗണനയിലിരിക്കുന്ന വേറൊരു അക്ഷരം മലയാളം രേഫമാണ്. സൂര്യൻ, ഭാര്യ എന്നൊക്കെ പണ്ട് യയുടെ മുകളിൽ കുത്തിട്ട് എഴുതുന്ന സമ്പ്രദായമുണ്ടായിരുന്നല്ലോ, ആ കുത്താണ് രേഫം എന്നതുകൊണ്ട് ഉദ്ദേശിക്കുന്നത്. ഒറ്റ നോട്ടത്തിൽ ഈ അക്ഷരത്തിന് മലയാളം യൂണിക്കോഡിൽ ഈ അക്ഷരം വന്നാൽ പ്രശ്നമുണ്ടെന്ന് ആർക്കും തോന്നില്ല. പക്ഷേ സൂര്യൻ, ഭാര്യ, സർക്കാർ എന്നൊക്കെ രണ്ടു രീതിയിൽ എഴുതിത്തുടങ്ങുകയും സാങ്കേതികമായി ആ വാക്കുകൾ വ്യത്യസ്തങ്ങളാവുകയും ചെയ്യുന്ന പ്രശ്നമാണ് ഉണ്ടാവാൻ പോകുന്നത്.

കാലം തെറ്റിയ പരിഷ്കാരങ്ങളും മലയാളഭാഷാവിദഗ്ധരുടെ പ്രാതിനിധ്യമില്ലാത്ത തീരുമാനങ്ങളുമാണ് ഇന്ന് യൂണിക്കോഡിൽ ഉണ്ടാവുന്നത്. ഈ തീരുമാനങ്ങൾ ഭാഷയ്ക്കു വേണ്ടിയല്ലാതെ സോഫ്റ്റ്‌വെയർ കമ്പനികളുടെ സോഫ്റ്റ്‌വെയറുകൾക്ക് വേണ്ടിയാവുമ്പോൾ ദിശ നഷ്ടപ്പെടുന്നു. ഭാരത സർക്കാറിനും തമിഴ്നാട് ഗവൺമെന്റിനും യൂണിക്കോഡ് കൺസോർഷ്യത്തിൽ വോട്ടവകാശമുള്ള അംഗത്വമുണ്ട്. പക്ഷേ അത് ഭാഷ ഉപയോഗിക്കുന്നവരുമായി ചർച്ചചെയ്തോ ചുരുങ്ങിയത് അവരെ അറിയിച്ചോ ഒന്നുമല്ല ഉപയോഗിക്കപ്പെടുന്നത്. മലയാളഭാഷയുടെ ഡിജിറ്റൽ ഭാവിയിലെ സുപ്രധാന തീരുമാനങ്ങൾ വേറെ എവിടെയോ വെച്ച് ആരും അറിയാതെ എടുക്കുമ്പോൾ ആ തീരുമാനങ്ങളിൽ നമ്മുടെ ശബ്ദം കേൾപ്പിക്കേണ്ടതില്ലേ? ■

▫ വി.കെ.ആദർശ്

വിക്കിപീഡിയയും മലയാളവും

പതിനായിരം ലേഖനങ്ങളുമായി വിക്കിപീഡിയ മലയാളം പതിപ്പ് മുന്നേറുന്നു

ഇന്റർനെറ്റിന്റെ കടന്നുവരവോടെ അറിവിന്റെ ഉൽപാദനത്തിനും വിതരണത്തിനും മുമ്പെങ്ങുമില്ലാത്ത മാറ്റവും ഗതിവേഗവുമാണ് സംഭവിച്ചുകൊണ്ടിരിക്കുന്നത്. കംപ്യൂട്ടർ, മൊബൈൽ ഫോൺ, കൊണ്ടുനടക്കാവുന്ന പിഡിഎ (PDA- Personal Digital Assistant) പോലെയുള്ള ഉപകരണങ്ങളുടെ വ്യാപനവും അതിന്റെ ഫലമായുണ്ടായ താരതമ്യേന താങ്ങാവുന്ന മുതൽമുടക്കും ആകണം ഇത്തരം ഉപകരണങ്ങൾ ജനകീയമാകാൻ കാരണം. ഇതിനോടൊപ്പം തന്നെ ഉയർന്ന വിവര കൈമാറ്റശേഷിയുള്ള ബ്രോഡ് ബാൻഡ് ഇന്റർനെറ്റിന്റെ ലഭ്യതയും വിവരവിപ്ലവം താഴെത്തട്ടിൽവരെയെത്തുന്നതിന് സഹായിച്ചു. ആദ്യകാലത്ത് വിവരം തേടാനുള്ള കേവലോപാധി എന്ന നിലയിലായിരുന്നു ഇന്റർനെറ്റിനെ കണ്ടതും പരിഗണിച്ചതും. എന്നാൽ രണ്ടാം തലമുറ (വെബ് 2.0) ഇന്റർനെറ്റ് യുഗത്തിൽ വിവരം പങ്കുവയ്ക്കുന്നതിനും അതുവഴി രൂപംകൊള്ളുന്ന വിജ്ഞാന നിർമിതിയിൽ ഉപയോക്താവ് കൂടി പങ്കാളിയാകുന്ന കാഴ്ചയാണ് കാണുന്നത്. ബ്ലോഗ്, വീഡിയോ ഷെയറിംഗ് സേവനങ്ങൾ, ഓർക്കുട്ട്, ഫേസ്ബുക്ക് പോലുള്ള വ്യക്തിഗത സാമൂഹിക ശൃംഖല, സംവിധാനങ്ങൾ, വിക്കിപീഡിയ എന്നിവയെല്ലാം വെബ് 2.0-യിലെ ശക്തമായ കണ്ണികളാണ്.

ഇന്റർനെറ്റിന്റെ പ്രാരംഭകാലത്തുണ്ടായിരുന്ന ഏറ്റവും വലിയ ആശങ്ക ഭാഷയുമായി ബന്ധപ്പെട്ടതായിരുന്നു. 90കളുടെ തുടക്കത്തിൽ സാമൂഹികശാസ്ത്രജ്ഞരും ഭാഷാസ്നേഹികളും ഭയപ്പെട്ടിരുന്നത് "ഇന്റർനെറ്റ് + ഇംഗ്ലീഷ്" എന്ന കൂട്ടുകെട്ടിനെയായിരുന്നു. അക്കാലത്തെ സാങ്കേതിക സ്ഥിതി വച്ച് ഈ ആശങ്ക ശരിയായിരുന്നുതാനും. ഇന്റർനെറ്റിൽ അല്ലെങ്കിൽ കംപ്യൂട്ടറിന് വഴങ്ങുന്ന ഭാഷ ഇംഗ്ലീഷ് മാത്രമാണെന്നും സമീപഭാവിയിൽ തന്നെ ഇന്റർനെറ്റിന്റെ ചിറകിലേറി ഇംഗ്ലീഷ് അധീശത്വം സ്ഥാപിക്കുമെന്നതുമായിരുന്നു അന്നത്തെ ആശങ്കയുടെ കാതൽ. എന്നാൽ പിന്നീട് പ്രാദേശിക ഭാഷാ ഫോണ്ടുകളുടെയും ഇന്റർനെറ്റിൽ ഇതിനെ തുണയ്ക്കുന്ന യൂണികോഡ് വകഭേദങ്ങളും എത്തിയതോടെ ബ്ലോഗ്, സോഷ്യൽ നെറ്റ്‌വർക്കിങ് ഇടങ്ങൾ, വിക്കിപീഡിയ എന്നിവയിലൂടെ മലയാളം അടക്കമുള്ള ചെറുഭാഷകൾ അതിന്റേതായ ശക്തവും വ്യക്തവുമായ സാന്നിധ്യം അറിയിക്കാൻ തുടങ്ങി. ഏതാണ്ട് ഈ കാലത്ത് തന്നെയാണ് ഒന്നാം തലമുറ വെബ്ബിൽ നിന്ന് രണ്ടാം തലമുറയിലേക്കുള്ള മാറ്റമെന്നതും ഏറെ ശ്രദ്ധേയമാണ്. സഹവർത്തിത്തം (collaboration) അറിവിന്റെ സാമൂഹിക കൊടുക്കൽ വാങ്ങൽ (knowledge based social interaction) സക്രിയ പങ്കാളിത്തം (Active participation), മറ്റ് ഉപകരണങ്ങളിലൂടെയുള്ള വിവരകൈമാറ്റം (communication through Allied ICT products) എന്നിവയാണ് വെബ്ബ് 2.0ന്റെ മുഖമുദ്ര.

വിക്കിപീഡിയ:

വിജ്ഞാനത്തിന്റെ ശേഖരമാണല്ലോ എൻസൈക്ലോപീഡിയകൾ. ബ്രിട്ടാണിക്ക