யுனிக்கோடு இட ஒதுக்கீட்டில் தமிழுக்கு அநீதி

இரா. துரைப்பாண்டி

கணினி உலகில் அதிகம் தேவைப்படும் ஒரு பொருள் எழுத்துரு. ஆனால் பொதுவான அமைப்புகள் ஏதுமின்றி ஒவ்வொருவரும் ஆர்வம் காரணமாகவும் வசதிகள் காரணமாகவும் தமக்கேற்றவகையில் எழுத்துக்களின் சீர்மையை வடிவமைத்தால் ஒருவர் பயன்படுத்தும் எழுத்துருவில் அடங்கிய கோப்பை மற்றொருவர் பயன்படுத்த இயலாது போகிறது. இதற்கு தீர்வு காணும் வகையில் தற்போது ஒரே வகையிலான தரச்சீர்மையுடைய தமிழ் எழுத்துருவை உருவாக்கும் பணி நடந்துவருகிறது. அதிலும் பல்வேறு குழறுபடிகள் உள்ளன. இவை குறித்து நம்முடன் உரையாடுகிறார் தமிழ்க் கணினி வல்லுநர் இரா.துரைப்பாண்டி. இவர் தமிழ் எழுத்துருக்களையும் விசைப்பலகை இயக்கிகளையும் இலவசமாக வழங்கிவருவதோடு, எளிய ஆங்கிலச் சொற்றொடர்களை தமிழுக்கு மொழிபெயர்ப்பு செய்யும் தமிழ்ப்பொறி என்ற மென்பொருளை தமிழ் இணையப் பல்கலைக்கழகத்தின் நிதியுதவியுடன் உருவாக்கி அது தமிழக முதல்வரால் 20.08.2003 அன்று சென்னையில் வெளியிடப்பட்டுள்ளது. துரைப்பாண்டியைச் சந்தித்து உரையாடியதிலிருந்து சில பகுதிகள் :

கணினியில் தமிழ் எவ்வாறு பயன்படுத்தப்படுகிறது?

அடிப்படையில் கணினிக்கு எந்தவொரு மொழியோ எழுத்தோ தெரியாது. அதற்குத் தெரிந்ததெல்லாம் சுழியும் ஒன்றும் (0,1) தான். எண்களை எப்படி இரண்டு அடிமான எண்களாக எழுதுவது என்று (15 = 1111, 1345 = 101010000010 நாம் பள்ளியில் படித்திருக்கிறோம் அல்லவா, அதுபோலவே நாம் கணினியில் பய்னபடுத்தும் எண்களையும் எழுத்துகளையும் கணினி தனக்குத் தெரிந்த இரண்டு அடிமான எண்களாக மாற்றிக்கொள்ளும்.

கணினியில் இந்த வேலையைத் தரக்குறியேற்றம் செய்கிறது. ஒவ்வொரு எழுத்தையும் அதற்கென குறிக்கப்பட்ட ஒரு எண்ணாக மாற்றி அந்த எண்ணையே கணினி கையாளுகிறது. அதாவது, எழுத்து மட்டுமல்லாமல் அரைப்புள்ளி, காற்புள்ளி போன்ற இன்னும் பல குறியீடு களையும் அது எண்களாகவே சேமிக்கிறது. நாம் அளிக்கும் 0 முதல் 9 வரையான எண்களையேகூட அது வேறு எண்களாக மாற்றித்தான் சேமிக்கிறது.

இதை இன்னொரு இடத்துக்கு அனுப்பினால் அதைப் பெற்றுக்கொண்ட கணினி அதேபோல மறுபடியும் எழுத்துக்களாக மாற்றுவதற்கு இதேவகையாக தரக்குறியேற்றமுடைய எழுத்துரு தேவைப்படுகிறது. அப்போதுதான் தரநிர்ணயத்தின் தேவை பிறக்கிறது.

இந்த வகைத் தரநிர்ணயத்தில் முதலில் நந்தது ஆஸ்கி. இந்த தரநிர்ணயத்தில் ஆங்கில எழுத்துக்களுக்கு மட்டுமே இடம் உண்டு. வெறும் 26 எழுத்துகளே உடைய ஆங்கில எழுத்துக்களுக்கும் மற்றும் எண்களுக்கும், குறியீடுகளுக்குமாய்ச் சேர்த்து 128 இடங்கள் இந்த ஆங்கி தரநிர்ணயத்தில் ஒதுக்கப்பட்டுள்ளன. தமிழில் சிலர் எழுத்துருக்களை வடிவமைத்த சிலர் ஆஸ்கியின் முதல் 128 இடத்திற்குள்ளும்கூட நுழைந்து தமிழ் எழுத்துக்களை அமைத்தனர். இதனால் பெரும் சிக்கல்கள் எழுந்தன.

8பிட்டுகள் உள்ள ஒரு பைட்டில் 256 எழுத்துக்களை குறியிடுவதற்குச் சாத்தியம் இருக்கிறது. ஆங்கிலத்திற்கான 128 போக எஞ்சிய 128 இடங்களில் இந்திய மொழிகள் அனைத்தையும் இடம்பெறச் செய்யும் அளவில் இந்திய அரசின் மின்னணுவியல் துறை இஸ்கி என்ற குறியீட்டு முறையை வடிவமைத்தது. எனவே இந்திய மொழிகள் அனைத்திற்கும் தனித்தனியாக 128 இடங்கள் கொண்ட வடிவமைப்பு நடைமுறைக்கு வந்தது. இதே முறையில்தான் தமிழைப் பயன்படுத்தும் முறையில் வெளிநாட்டுத் தமிழர்கள் தனியாக திஸ்கி என்ற முறையில் எழுத்துருக்களை வடிவமைத்தனர்.

கடைசியாக, சென்னையில் 1999ஆம் ஆண்டு நடைபெற்ற தமிழ் இணைய மாநாட்டில் தாப், தாம் ஆகிய இரு வடிவமைப்புகள் தமிழக அரசால் ஏற்றுக்கொள்ளப்பட்டன. அதில் தாம் என்பது தமிழ் மட்டுமேயான வடிவமைப்பு, தாப் என்பது தமிழும் ஆங்கிலமும் இணைந்த வடிவமைப்பு. 2001 தமிழ் இணைய மாநாட்டில் தாப்-பா திஸ்கி-யா எது பொதுக்குறியீடு என்ற கேள்வி எழுந்தபோது இரண்டுக்கும் பொதுவான மென்பொருள் மாற்றிகளைக் கொண்டு இரண்டையும் பயன்படுத்திக்கொள்ளலாம் என பரிந்துரைத்தது நான்தான். அதுவே ஏற்றுக்கொள்ளப்பட்டது. இந்த எழுத்துருக்கள் ஒன்று அல்லது இரண்டு மொழிகள் மட்டுமே அடங்கியவை. இதன் அடுத்தகட்டமாக உலக மொழிகள் பெரும்பான்மை எல்லாம் அடங்கிய, தரப்படுத்தப்பட்ட எழுத்துருவை உருவாக்கும் முயற்சி எழுந்தது. அதுவே யுனிக்கோடு

அதுகுறித்து சொல்லுங்ளேன்?

யுனிக்கோடு கன்சார்டியம் அமைப்பு யுனிக்கோடு என்ற புதிய தரக்குறியீட்டு முறையை வடிவமைத்துள்ளது. அதாவது, முந்தைய 8 பிட் எழுத்துருவில் மொத்தம் 256 இடங்கள்தான் உள்ளன. ஆனால் இப்போதைய 16 பிட் எழுத்துருவில் மொத்தம் 65,000 எழுத்துகளுக்கு இடம் இருக்கிறது. எனவே, இதில் முறையாக இடஒதுக்கீடு செய்தால் உலகமொழிகள் அனைத்திற்கும் ஒந்த ஒரு வடிவமைப்பே தாராளமாகப் போதும்.

அதன்படி இதில் சேர்ப்பதற்காக ஒவ்வொரு மெமாழியினதும் எழுத்துகளின் பட்டியலைக் கேட்டது யுனிக்கோடு கன்சார்டியம். ஆனால், இங்கேதான் விதி விளையாடியது அல்லது தூங்கப்போனது, அதாவது, சோம்பேறித்தனத்திற்கு அல்லது பொறுப்பற்றத்தனத்திற்குப் பெயர்பெற்ற நம் அதிகாரவர்க்கம் பழைய இஸ்கி அடிப்படையிலான 128 எழுத்துகளின் பட்டியலை அப்படியே யுனிக்கோடு கன்சார்டியத்திற்குப் ஃபார்வர்ட் செய்தது. தமிழைப் பொருத்தவரை அதிலும் 67 காலியிடங்கள் அதுதான் தற்போது ஏற்றுக்கொள்ளப்பட்டு நடைமுறையில் உள்ளது. இந்த அறுபத்தைந்தாயிரத்தில்2944லிருந்து 3071 வரையிலான 128 இடங்கள் தமிழுக்காக ஒதுக்கப்பட்டன. இங்கேதான் சிக்கல் தோன்றியது

அப்படி என்ன இதில் சிக்கல் இருக்கிறது?

தமிழில் 247 எழுத்துக்கள் என்பது தெரியும். இதையே மெய்யெழுத்து 18, உயிரெழுத்து 12 ஆய்தம் 1 ஆக 31 என்றும் சொல்லாம். யுனிக்கோடில் கிட்டத்தட்ட இந்த முறைதான் கையாளப்பட்டிருக்கிறது, மேலும் இதன் அமைப்பு அகர வரிசையில் இல்லை.

அதாவது தமிழ் எழுத்துகள் அனைத்தும் யுனிக்கோடு எழுத்துருவில் இடம்பெற்றிருந்தால் சிக்கல் இல்லை. புதியபார்வை என்பதைத் தட்டச்சு செய்தால் 8 எழுத்துகள்தான் ஆகும். ஆனால் இப்போதைய யுனிக்கோடு முறையில் அதற்கு வழியின்றி புதியபார்வை என்பது ப+உ, த+இ ய ப+ா, ர் வ+ை என்று 11 எழுத்துகளிலேயே பதிவாகும். எனவே, எழுத்துகளின் எண்ணிக்கை கூடுவதோடு, கோப்பின் அளவும் கூடுகிறது. எனவே, இணையத்தில் அனுப்பும் நேரம், இறக்கும் நேரம், அகரவரிசைப்படுத்தல், தேடல் போன்றவற்றிற்கான நேரங்கள் அதிகரிக்கின்றன. இன்றைய தகவல் தொழில்நுட்ப யுகத்தில் தகவல் என்பது சக்தியுடையது என்கிறார்கள். ஆனால் இப்படி தமிழ் ஒரு கால் உடைக்கப்பட்ட நிலையில் இருக்குமாயின் நம்மால் ஆங்கிலத்தோடு போட்டிபோட முடியாது. அதாவது ஒரு தகவல் ஆங்கிலத்தில் செல்லும் அதேநேரம் நமக்கு கிடைப்பதற்குப் பதிலாக தாமதப்படும். நாளைய போட்டி நிறைந்த உலகில் தகவல் கிடைக்கும் நேரத்தைப் பொறுத்து முடிவெடுக்கும் ஆற்றல் வெற்றிகளுக்குக் காரணமாக இருக்கும் எனும் நிலையில் இந்த ஊனமுற்ற தொடர்புமுறை வளர்ச்சியை நோக்கமாகக் கொண்தல்ல. இதனால் கணினித்தமிழ் வளர்ச்சியும் பாதிக்கப்படும். நாம் இன்றைய உடனடி ஆதாயங்களை மட்டுமே கருத்தில் கொண்டு திட்டமிடுவோமேயானால் நமது சந்ததியினருக்குக் கிடைக்கவேண்டிய நாளைய வாய்ப்பைப் பறித்தவர்களாவோம்.

இது குறித்து எங்காவதுப் பேசப்பட்டுள்ளதா?

1999இல் சென்னையில் நடைபெற்ற தமிழ் இணைய மாநாட்டில் யுனிக்கோடில் தமிழுக்கான இடம் போதாது என்று குரல் எழுப்பப்பட்டது. கூடுதல் இடம் கேட்கவும், இதனால் ஏற்பட்டுள்ள நிலை குறித்து ஆராயவும் ஒரு பணிக்குழு அமைக்கப்பட்டது. தமிழ் இணையப் பல்கலைக் கழகத்தில் நடத்தப்பட்ட ஆய்வில் தற்போதைய எழுத்துரு அமைப்பால் கோப்பில் 30% - 40% அதிக இடமும் தாமதமும் ஏற்படுகிறது என்று ஆய்வு முடிவுகள் நிரூபித்திருக்கின்றன. கணினியின் வேகம் எந்தளவுக்கு அதிகரித்தாலும் நமக்கு இந்த இடைவெளி இருக்கவே செய்யும்.

இதுபோன்ற இணையத் தமிழ்ப் பிரச்சினைகளைத் தீர்க்கும் நோக்குடன் உலகளவில் ஒருங்கிணைப்பை ஏற்படுத்துவதற்காக உத்தமம் (உலகத் தமிழ் தகவல் தொழில்நுட்ப மன்றம்) என்ற அமைப்பு இயங்கி வருகிறது.

யுனிக்கோடு கன்சார்டியத்தில் இந்திய அரசு, தமிழக அரசு, உத்தமம் ஆகியோர் உறுப்பினர்களாக உள்ளனர். ஆனால் யுனிக்கோடு கன்சார்டிய கூட்டங்கள் எதிலும் இந்திய அரசின் பிரதிநிதிகள் யாரும் இதற்காகக் குரல்கொடுக்கவேயில்லை. ஆனந்தகிருஷ்ணன் அவர்கள் உத்தமத்தின் தலைவராக இருந்ததுவரை இந்த முறை மாற்றப்பட வேண்டும் என்ற கருத்து வலிமையுடன் இருந்து வந்தது. ஆனால் இப்போதைய தலைவரான முத்துநெடுமாறன் போன்றோருக்கு அதில் அந்த அளவு ஈடுபாடு இல்லை. இப்போது இருப்பதேபோதும் என்ற மனநிலையிலேயே இவர்களைப் போன்றவர்கள் இருக்கின்றனர். எதிர்பாராதவிதமாக தாம் எப்போதோ சொல்லிவிட்ட ஒரு கருத்தை இன்றை நிலையில் மாற்றிக் கொள்வதற்கு அவர்கள் தயாராக இல்லை. மாறாக, பழையதை நியாயப் படுத்தும் நிலையில் இருக்கிறார்கள். இது மாற்றத்தை ஏற்காத பழமைவாதப் போக்காக இருக்கிறது.

மைக்ரோசாஃப்ட் போன்ற பெரிய நிறுவனங்கள் எல்லாம் ஏற்றுக்கொண்ட பின்னர் அதை மாற்றுவது சுலபமா?

சீன, கொரிய மற்றும் ஜப்பான் எழுத்துகள் சித்திர எழுத்துகள் எண்ணிக்கையில் தமிழைவிடப் பல மடங்கு அதிகமானவை யுனிக் கோடில் கொரிய மொழிக்கு மட்டும் 12,177 இடங்கள் ஒதுக்கப்பட்டுள்ளன. சீன, கொரிய மற்றும் ஜப்பான் மொழிகளுக்குச் சேர்த்து சுமார் 25,000 இடங்கள் வரை ஒதுக்கப்பட்டிருக்கிறது. சிங்களர்கள் கூட 400 இடங்களை வாங்கிவிட்டார்கள். ஆக, செயலற்றவர்கள் தமிழர்களும் இந்தியர் களும்தான்.

தாங்கள் கேட்ட இடத்தை யுனிகோடு தரமறுத்ததால் சீனா அதிரடியாக தங்களுக்கென்று தனியே ஒரு குறியீட்டு முறையை வகுப்பதாகவும் அந்தக் குறியீட்டு முறையிலான மென்பொருள்களையே இனி சீனாவுக்குள் பன்னாட்டு நிறுவனங்கள் விற்கமுடியும் என்றும் அறிவித்தது. அவ்வளவுதான் எல்லா வணிகர்களும் அதற்குப் பணிய வேண்டிவந்தது. நம்முன் உள்ள வரலாறு.

தமிழுக்கு எத்தனை இடங்கள்தான் வேண்டும்?

தமிழ் எழுத்துக்கள், சிறப்புக் குறியீடுகள் போன்றவற்றைப் பயன்படுத்துவதற்காக குறைந்தபட்சம் 388 இடங்கள் தேவை. எனவே இந்திய மொழிகள் அனைத்திற்கும் தனித்தனியாக 512 இடங்களுடன் கூடிய புதிய இடஒதுக்கீடு தேவை. அப்படி ஒதுக்கீடு கிடைக்கவில்லை யெனில் இந்தியாவின் எட்டாம் அட்டவணைக்கு உட்பட்ட அனைத்து மொழிகளுக்குமான தனி 16 பிட் தரக்குறி வரிசையை இந்திய அரசு வெளியிட வேண்டும். அனைத்து பன்னாட்டு நிறுவனங்களும் அதன் அடிப்படையில் மென்பொருட்களை வடிவமைக்க வேண்டும் என்று நிர்ணயம் செய்யலாம். ஆக மொத்தத்தில் கணினித் துறையில் பன்பாட்டு நிறுவனங்களின் பிடியில் தமிழும் ஏனைய இந்திய மொழிகளும் இருக்கின்றன. ஓர் இடஒதுக்கீட்டுப் போராட்டமாக நாடு தழுவிய அளவில் இது மாறி வெற்றியடையவேண்டும்.

மைக்ரோசாஃப்பட், சன் மைக்ரோ சிஸ்டம் போன்ற நிறுவனங்கள் தற்போதுள்ள கட்டமைப்பை ஒட்டியே தமது இயக்கத் தொகுப்புகளை வெளியிட்டிருப்பதால், செலவு பிடிக்கும் புதிய வடிவமைப்பைத் தவிர்க்கவே முயற்சிப்பார்கள். அதற்கு நேரடி உதாரணம் உத்தமம் அமைப்பில் உறுப்பினராக உள்ள மைக்கேல்கெப்ளான் என்பவர் மைக்ரோசாஃப்பட் நிறுவனத்தைச் சார்ந்த அமெரிக்கர். அவருக்கு உத்தமத்தில் என்ன வேலை என்பது அடுத்த கேள்வி.

யுனிக்கோடு அமைப்பிற்கு எதிராக சர்வதேச மன்றங்களில் வழக்குத் தொடரவேண்டும் என்கிற ஆலோசனை ஓர் உத்தமம் உறுப்பினரால் வழங்கப்பட்டபோது உடனடியாக கடும் எதிர் வினையாற்றியவர் மைக்கேல் கெப்ளான். ஏனெனில் இந்திய மொழிகள் அனைத்திலும் யுனிக்கோடு தொடர்பான பெரும் விவாதம் நிகழ்வது தமிழ்ச் சூழலில் மட்டும்தான். அதனால்தான் திட்டமிட்டே கருத்துக்களைத் தீர்மானிக்க வேண்டிய குழுக்களில் அவர் உடனடியாக இடம் பிடித்திருக்கிறார். சிக்கல் புரிகிறதா உங்களுக்கு?

ஆனால், போதுமான இடங்களைப் பெறுவதற்காக நாம் போராடியே ஆகவேண்டும். ஏனென்றால் இது நமது எதிர்காலம் தொடர்பான பிரச்சினை. எதிர்காலத்தில் அனைத்துப் பணிகளும் கணினிமயமாக்கப்டும்போதுதான் நமக்கு இதன் முக்கியத்துவம் புரியும். அறிவியலும் தொழில் நுட்பமும் மக்களுக்கானவையாகத்தான் இருக்க வேண்டுமேயல்லாது நிறுவனங்களுக்காக இருத்தலாகாது.

சந்திப்பு : அசுரன்

(புதிய பார்வை அக்டோபர் 1, 2004)