



ইউনিকোড, ইন্টারনেটে বাংলা হরফের জটিলতা

মোস্তাফা জব্বার

ডাক ও টেলিযোগাযোগ মন্ত্রী

ইউনিকোড, ইন্টারনেট এবং বাংলা হরফের বিষয়গুলো আলোচনা করতে হলে আমাদের অক্ষরমালা এখন কোন অবস্থাতে বিরাজ করে তার বিবরণটি জানা দরকার। বাস্তবতা হলো, বাংলা ভাষার নামে প্রতিষ্ঠিত বিশ্বের একমাত্র রাষ্ট্র বাংলাদেশের বাংলা অক্ষরমালা ইউনিকোডের ক্ষমাহীন একতরফা চাপিয়ে দেয়া সিদ্ধান্তের জন্য চরম বিপন্নতায় রয়েছে। ইউনিকোড বাংলা হরফকে যথাযথভাবে এনকোডিং না করার ফলে ইন্টারনেটে অন্তত তিনটি বাংলা হরফ ব্যবহারের চরম জটিলতা বিরাজ করছে। আমরা যদি জাতিগতভাবে রাষ্ট্রীয় উদ্যোগে এই সমস্যার সমাধান করতে না পারি, তবে সেটি আমাদের বর্তমান ও ভবিষ্যতের ভাষা ব্যবহারের জন্য চরম সংকট তৈরি করবে। এমনকি ড, ঢ, য় দিয়ে ইন্টারনেটের ডোমেইন নামে কোনো শব্দ লেখার সুযোগ নাও থাকতে পারে। ২০১০ সাল পর্যন্ত ইউনিকোড কর্তৃপক্ষের সাথে কোনো যোগাযোগ না করা ও তার পরেও বাংলা হরফের এনকোডিং নিয়ে ইউনিকোড কর্তৃপক্ষের সাথে যথাযথ যোগাযোগ রক্ষা না করায় আমরা এমন দুর্গতিতে পড়েছি। বর্তমান সংকটটি সম্পর্কে একটু ধারণা পাওয়ার জন্য আমরা একটি প্রতিবেদনের দিকে নজর দিতে পারি।

প্রথমেই আমি আমার স্নেহভাজন পল্লব মোহাইমেনের একটি লেখা তুলে ধরতে চাই। '১৯ সালের ফেব্রুয়ারি মাসে পল্লব নিবন্ধটি লেখে। পল্লবের লেখাটির শিরোনাম ছিল, 'ইউনিকোডে বাংলা রীতি মানা হয়নি'।

কমপিউটারে বাংলা হরফ ব্যবহারের জটিলতার চূড়ান্ত সমাধান পদে পদে আটকে থাকছে। আসকিভিত্তিক পুরনো ফন্ট স্মার্টফোন ও ইন্টারনেট প্রভৃতি ওয়েবক্ষেত্রে বাংলা বর্ণমালা ব্যবহার করা যেত না। ফন্টের সর্বজনীন কারিগরি ব্যবস্থা 'ইউনিকোড' আসার পর এ জটিলতা দূর হয়েছিল। তাতে সে সংকট দূর হলেও ইউনিকোডের ফন্ট বিন্যাসে বাংলা নিয়ে নতুন জটিলতা সৃষ্টি হয়েছে।

ড, ঢ, য় এই তিনটি বাংলা বর্ণ ইউনিকোডের সর্বশেষ ১১তম সংস্করণে সরাসরি লেখার ব্যবস্থা নেই। ইউনিকোড তৈরি ও ব্যবস্থাপনার আন্তর্জাতিক সংগঠন ইউনিকোড কনসোর্টিয়াম ইউনিকোডে বাংলা ভাষার সাংকেতিক ব্যবস্থা বা কোডসেট রেখেছে ভারতের হিন্দি লিপি দেবনাগরীর অনুকরণে। এ কারণে বর্ণগুলো

	098	099	09A	09B	09C	09D	09E	09F
0	০	ঐ	ঔ	ৱ	ি		ঋ	ূ
1	ৃ		ড		ৄ		ঌ	৆
2	৅		ঢ	ল	৆		৏	ৈ
3	ে	ও	ণ		৏		ঔ	ট
4		ঔ	ত		৐			৊
5	অ	ক	খ					ো
6	আ	খ	দ	শ			ৌ	্
7	ই	গ	ধ	ষ	়ে	়ী	ৎ	ৎ
8	ঈ	য	ন	স	়ৈ		৐	৐
9	ঊ	ঙ		হ			৑	৑
A	ঋ	চ	প				৒	৒
B	ঌ	ছ	ফ		়ৌ		৓	৓
C	ৎ	জ	ব		়ৌ	ড়	৔	৔
D		ঝ	ভ	হ		ঢ	৕	৕
E		ঞ	ম	া	ং		ট	৖
F	এ	ট	য	ি		য়	ৗ	

লেখতে ড, ঢ ও য-এর নিচে নোকতা অর্থাৎ বাড়তি একটা ডট (.) দিতে হবে। এটি ভবিষ্যতে ইন্টারনেটে বাংলা লেখার ক্ষেত্রে সীমাবদ্ধতা তৈরি করবে বলে ভাষা বিশেষজ্ঞেরা মনে করছেন।

সর্বশেষ হালনাগাদ ইউনিকোডে আরও কিছু অব্যবস্থা বিশেষজ্ঞেরা চিহ্নিত করেছেন। বাংলা লেখার স্বাভাবিক প্রকৃতি অনুযায়ী এ-কার, ই-কার, ঈ-কার বর্ণের আগে না করে পরে টাইপ

করতে হচ্ছে। বাংলা ভাষায় ব্যবহৃত যতিচিহ্ন দাঁড়ির বদলে এসেছে দেবনাগরী বর্ণমালার মোটা ও দীর্ঘ দাঁড়ি। এতে বাংলা ভাষার দ্বৈত দাঁড়ি রাখা হয়নি। বাংলাদেশি টাকার চিহ্নকে অভিহিত করা হয়েছে 'বেঙ্গলি রুপি' হিসেবে। ভাষাবিজ্ঞানী ও ভাষা প্রযুক্তিবিদেরা বলছেন, ইউনিকোডে বাংলা ভাষার প্রকৃত প্রতিফলন ঘটেনি। সংশ্লিষ্ট ব্যক্তিরা বলছেন, ইউনিকোডের বাংলা এর নিজস্ব ভাষারীতি অনুসারেই হওয়া উচিত।

ওয়েব ঠিকানা লিখতেও সমস্যা হবে

ইন্টারনেটের ডোমেইন ঠিকানা বরাদ্দকারী মার্কিন সংস্থা ইন্টারনেট করপোরেশন ফর অ্যাসাইনড নেমস অ্যান্ড নাম্বারস (আইসিএএনএন) ইংরেজি ছাড়াও অন্য ভাষায় ওয়েব ঠিকানা বরাদ্দ করে থাকে। বাংলায় ওয়েব ঠিকানার জন্য ডট বাংলাও করা হয়েছে। ইউনিকোডের বাংলা ছকে সরাসরি এই অক্ষরগুলো না থাকায় বাংলায় ওয়েব ঠিকানা লিখতে সমস্যা হবে।

বাংলাদেশ কমপিউটার কাউন্সিলের (বিসিসি) ভাষা-প্রযুক্তি বিশেষজ্ঞ এবং জাহাঙ্গীরনগর বিশ্ববিদ্যালয়ের সহযোগী অধ্যাপক মামুন অর রশীদ জানান, ভারতের উদ্যোগে সে দেশে হিন্দি ভাষার অনুসারে নয়টি আঞ্চলিক ভাষার জন্য ইউনিকোডে নিউ ব্রাহ্মী জেনারেশন প্যানেল তৈরি করা হয়েছে। এর মধ্যে হিন্দি ভাষার ব্রাহ্মীলিপি থেকে বাংলা লেখার ব্যবস্থা রাখা হয়েছে। ফলে নোকতা যোগ করে ড, ঢ ও য় লেখার ক্ষেত্র তৈরি হয়েছে।

মামুন অর রশীদ বলেন, 'বাংলা লিপির উদ্ভব দেবনাগরী থেকে নয়। বাংলা এভাবে থাকলে ইন্টারনেটে বাংলায় ডোমেইন তৈরি করা যেমন সমস্যা হবে, তেমনি সার্চ ইঞ্জিনে তথ্য খুঁজতেও জটিলতার মধ্যে পড়তে হবে। এর সমাধান করতে হলে ইউনিকোড কনসোর্টিয়াম ও আইসিএএনএন দুই জায়গাতেই আমাদের কথা বলতে হবে।'

এখন ইউনিকোডে যেভাবে বাংলা রয়েছে, সেটা বাংলা হয়নি বলে মনে করেন ভাষাবিজ্ঞানী এবং জাহাঙ্গীরনগর বিশ্ববিদ্যালয়ের সাবেক অধ্যাপক মহাম্মদ দানীউল হক। তিনি বলেন, হিন্দিতে বা অন্য ভাষায় নোকতা দরকার হলেও বাংলায় দরকার নেই। ইউনিকোড ইন্টারফেসে প্রচুর জায়গা আছে। বাংলা অক্ষরগুলো সেখানে সরাসরি দিয়ে দিলেই হয়।

ইউনিকোডের বাংলা বাংলাদেশের মতের ▶

ইউনিকোড ও আইসিএএনএনের দাবি

এ বিষয়ে ডাক, টেলিযোগাযোগ ও তথ্যপ্রযুক্তিমন্ত্রী মোস্তাফা জব্বার বলেন, বাংলা ভাষা নিয়ে ইউনিকোডের বর্তমান ছক একেবারেই গ্রহণযোগ্য নয়। হিন্দিতে যা-ই থাকুক, বাংলায় তার অনুকরণ গ্রহণযোগ্য হতে পারে না। বাংলা স্বতন্ত্র ভাষা-আমাদের অক্ষরমালা ভিন্ন। কিন্তু ইউনিকোডে তার প্রতিফলন নেই। বাংলাদেশের অংশগ্রহণ না থাকায় ভারতের সাথে যোগাযোগ করেই ইউনিকোডের বাংলা তৈরি করা হয়েছে।

বিষয়টি সমাধানে সরকার একাধিক কমিটি গঠন করেছে। বিসিসি ও বিটিআরসি আইসিএএনএন ও ইউনিকোড কনসোর্টিয়ামের সভাগুলোতে ইউনিকোডের বাংলা বাংলাদেশের দাবি জোরালোভাবে তুলে ধরেছে। কিছুদিন পর জাপানে অনুষ্ঠিত আইসিএএনএনের সভায় বিটিআরসি বাংলাদেশের দাবি তুলে ধরবে বলে জানা গেছে।

পল্লবের এই প্রতিবেদনটি '১৯ সালের ফেব্রুয়ারি পর্যন্ত বাংলা ভাষার সাথে ইউনিকোড এবং আইকানের সাথে বিদ্যমান সংকট নিয়ে আলোচিত হয়েছে। এই নিবন্ধটি লেখার সময় পর্যন্ত সেই ক্ষেত্রে আরও কিছু অগ্রগতি রয়েছে। বস্তুত প্রচণ্ড চাপ সৃষ্টি করে আমি বিটিআরসি, কমপিউটার কাউন্সিল, ইন্টারনেট গভর্ন্যান্স ফোরাম ও অন্যান্য বিশেষজ্ঞদের নিয়ে বাংলা বর্ণমালার ইন্টারনেট এনকোডিং ও ইন্টারনেটে এর ব্যবহারের সংকটগুলো নিয়ে একটি ঐকমত্য তৈরি করতে পারি। পল্লব মোহাইমেনের প্রতিবেদনে যথার্থই উল্লেখ করা হয় যে আমরা আইকানের একতরফা আচরণে ক্ষুব্ধ হয়েছি। তারা দেবনাগরীর অনুসরণে আমাদেরকে নোকতা দিয়ে ড, ঢ, য লেখতে বাধ্য করার প্রক্রিয়ায় যাওয়ায় আমরা তীব্র প্রতিবাদ করি এবং একের পর সভায় যোগ দিয়ে আমাদের দাবি তুলে ধরি। এমনকি ১৯১৮ সালে স্পেনে আমি আইকানের প্রধান নির্বাহীর সাথেও কথা বলি। তবে সংকটটা হচ্ছে ভারত একতরফাভাবে দেবনাগরীকে তাদের সব ভাষার ওপর চাপিয়ে দেয়ার মানসিকতা থেকে নিউ ব্রাহ্মী লিপি নামক এক প্রচেষ্টার মধ্য দিয়ে বাংলা তিনটি অক্ষরকে দেবনাগরীর মতো নোকতামুক্ত করার প্রয়াস চালায়। নিউ ব্রাহ্মী জেনারেশন এই বিষয়ে ঢাকার বাংলা একাডেমিতে একটি সেমিনারের আয়োজন করে যাতে বাংলাদেশ ও পশ্চিমবঙ্গের বিশেষজ্ঞরা স্পষ্টভাবে ড, ঢ, য অক্ষরগুলো স্বাধীন ও এককভাবে লেখার পক্ষ মতামত দেন। আমি বাংলাদেশ রাষ্ট্রের পক্ষ নোকতা দিয়ে তিনটি অক্ষর লেখার পদ্ধতি দিয়ে লেখা যে গ্রহণযোগ্য নয় তা স্পষ্টভাবে দৃঢ়তার সাথে জানিয়ে দিই। এরপর আইকানের সভায়, আইজিএফের সভায় বা ইউনিকোডের সভায় বাংলাদেশ তার নিজের অবস্থান স্পষ্ট করে তুলে ধরে। দুঃখজনকভাবে গত অক্টোবরে ইউনিকোড কনসোর্টিয়ামের একটি সভায় আমাদের কমপিউটার কাউন্সিল যোগ না দেয়ায় আমাদের দাবি পেশ করার একটি সুযোগ আমাদের হাতছাড়া হয়। তবে এটি আমরা দৃঢ়ভাবে নিশ্চিত করতে চাইছি যে বাংলা বর্ণমালার ইন্টারনেটে ব্যবহারের ক্ষেত্রে আমরা কোনোভাবেই এক চুল আপস করব না। বিষয়টি এখন বিটিআরসি, আইজিএফ ও বিটিআরসিসহ একটি বিশেষজ্ঞ দলের গভীর মনোযোগের বিষয় হিসেবে গণ্য হচ্ছে। বিটিআরসিকে এরই মাঝে আইজিএফের সদস্যপদ গ্রহণের নির্দেশ দেয়া হয়েছে। সেই নির্দেশনা অনুসারে নভেম্বরের শেষ দিকে জার্মানিতে অনুষ্ঠিত ইন্টারনেট গভর্ন্যান্স ফোরামের সভায় প্রথমবারের মতো বিটিআরসির প্রতিনিধি বাংলাদেশের একটি প্রতিনিধিদলের সাথে যোগ দেবে এবং বাংলা ভাষা ইন্টারনেটে ব্যবহার যাতে বাধাহীন হতে পারে তার উদ্যোগ নেবে।

ভিত্তিতে হওয়া বাঞ্ছনীয় বলে মত দিয়েছেন পশ্চিমবঙ্গের বাংলা ভাষাবিজ্ঞানীরাও। রবীন্দ্রভারতী বিশ্ববিদ্যালয়ের সাবেক উপাচার্য ও ভাষাবিজ্ঞানী পবিত্র সরকার প্রথম আলোকে বলেছেন, 'নোকতা দিয়ে ড, ঢ ও য লেখার ব্যবস্থা একেবারেই ঠিক নয়। আমরা যেভাবে লিখি, সেভাবেই পুরো অক্ষর তৈরি হতে হবে। বাংলা ও হিন্দি লিপি আলাদা। দৃষ্টি প্রায় সমবয়সী। বাংলা দেবনাগরীজাত নয়।'

ইউনিকোডে বাংলা ভাষার সংযুক্তির ব্যাপারে পশ্চিমবঙ্গ থেকে প্রাতিষ্ঠানিকভাবে তেমন কোনো উদ্যোগ নেয়া হয়নি। পবিত্র সরকার বলেন, 'পশ্চিমবঙ্গের বাংলা আকাদেমি ও ভাষাপ্রযুক্তি গবেষণা পরিষদ এ ব্যাপারে কোনো উদ্যোগ নেয়নি। এ ব্যাপারে আমরা বাংলাদেশের ভূমিকাকে পুরোপুরি সমর্থন করছি। বাংলা ভাষা ও লিপির জন্য বাংলাদেশের আত্মত্যাগের ইতিহাস রয়েছে।'

ইউনিকোড ও বাংলাদেশ

ইউনিকোডের শুরু ১৯৮৭ সালে অ্যাপল কমপিউটারের উদ্যোগে। পরে মাইক্রোসফটসহ বড় বড় তথ্যপ্রযুক্তি প্রতিষ্ঠান এ উদ্যোগে যুক্ত হয়ে ইউনিকোড কনসোর্টিয়াম গঠন করে। অ্যাপলের পরিবেশক হিসেবে ১৯৮৮ সালে

খাইল্যান্ডে অ্যাপলের একটি সম্মেলনে অংশ নেন আনন্দ কমপিউটার্সের প্রধান নির্বাহী এবং বর্তমানে ডাক ও টেলিযোগাযোগ ও তথ্যপ্রযুক্তিমন্ত্রী মোস্তাফা জব্বার। ইউনিকোডের বাংলায় সে সময় ড, ঢ, য ও ৭ এই চারটি বর্ণ ছিল না। মোস্তাফা জব্বার বিষয়টি নিয়ে প্রশ্ন তোলেন এবং এগুলো যুক্ত হয়।

ইউনিকোড কনসোর্টিয়াম গঠনের পর বিভিন্ন দেশ এর সদস্য হয়। বাংলাদেশ দীর্ঘদিন এর সদস্য ছিল না। ফলে ইউনিকোডে বাংলা ভাষার ছক কেমন হবে, বাংলাদেশের পক্ষ থেকে সে বিষয়ে কিছু বলা যায়নি। ২০১০ সালে বাংলাদেশ প্রতিনিধি হিসেবে বিসিসি ইউনিকোড কনসোর্টিয়ামের সদস্য হয়।

পরবর্তী সময়ে এসেছে ইন্টারনেটে অন্য ভাষার টপ লেভেল ডোমেইন (যেমন ডট বাংলা) নামের বিষয়টি। এখানেও বাংলাদেশের প্রতিনিধিত্ব ছিল না। ২০১৮ সালের মাঝামাঝি বাংলাদেশ টেলিযোগাযোগ নিয়ন্ত্রণ কমিশন আইসিএএনএনের সদস্য হয়।

বর্ণমালার এনকোডিং

প্রাসঙ্গিকভাবেই আমাদের বাংলা বর্ণের এনকোডিং প্রমিতকরণ ও তার চ্যালেঞ্জের সাথে

ইউনিকোড এবং তার সাথে প্রাসঙ্গিক বিষয়াদি নিয়ে কথা বলতে চাই। বাংলা বর্ণমালার বিশ্বজনীন ব্যবহারের ক্ষেত্রে এটি খুবই গুরুত্বপূর্ণ একটি বিষয়। ধীরে ধীরে বাংলার অন্যান্য প্রমিতকরণ প্রসঙ্গটিও আলোচিত হতে পারে।

শুরুতেই ডিজিটাল যন্ত্রে বর্ণের প্রমিতকরণ নিয়ে কথা বলা যায়। প্রথমত সাধারণ মানুষের বোঝার জন্য এটি বলা দরকার যে, কমপিউটার বস্তুত কোনো ভাষা বোঝে না। এটি যুগ্ম সংখ্যায় কাজ করে এবং এর ভাষা শুধু ০ ও ১। দুনিয়ার সব ভাষাই সংকেত আকারে ০ ও ১ হিসেবেই কমপিউটারে প্রক্রিয়াকরণ হয়। আমরা কমপিউটারের পর্দায় যা দেখি সেটি শুধু আমাদের দৃশ্যমানতা। যুগ্ম সংখ্যাকে কমপিউটারের প্রোগ্রামের সহায়তায় আমাদের ভাষায় রূপান্তর করে দেখানো হয়।

ভাষার জন্য কমপিউটারের প্রয়োজনে প্রথমে রোমান হরফকে ১২৮টি (০ থেকে ১২৭) কোডের মাঝে চিহ্নিত করা হয়। একে (লোয়ার) আসকি কোড বলা হতো। ১৯৬০ সালের অক্টোবরে এটি নিয়ে কাজ শুরু হলেও এর প্রথম সংস্করণ প্রকাশিত হয় ১৯৬৩ সালে। এক সময়ে সেটাকে ২৫৬টি কোডে পরিণত করা হয়। ১২৮-এর ওপরের কোডগুলোকে উচ্চতর আসকি (১২৮ থেকে ২৫৫) বলা হতো। আমেরিকানদের তৈরি করা সেই কোডই কমপিউটারে রোমান হরফের মান হিসেবে প্রচলিত ছিল। আসকি কোডে দুনিয়ার আর কোনো বর্ণমালার কোনো কোড ছিল না। কালক্রমে আসকিকে আরও সম্প্রসারিত করে এক্সটেন্ডেড আসকি কোড হিসেবে চিহ্নিত করা হয়। যার ফলে লাতিন সব ভাষার সব বর্ণকে তার মাঝে অন্তর্ভুক্ত করা হয়। দুনিয়ার আর কোনো ভাষার কোনো বর্ণকে এই এক্সটেন্ডেড কোডেও যুক্ত করা হয়নি।

আমরা ১৯৮৭ সালে প্রথম যখন কমপিউটারে বাংলা চালু করি তখন রোমান হরফের সেই কোডেই (২৫৬টি) বাংলা হরফের চেহারা বা ফন্ট বানিয়ে তাকে ব্যবহার করতাম। এখনও আমাদেরকে মুদ্রণ ও প্রকাশনার ক্ষেত্রে সেই আসকি কোডই ব্যবহার করতে হয়। এর প্রধানত কারণ হচ্ছে প্রকাশনার খুবই জনপ্রিয় অ্যাপ্লিকেশন প্রোগ্রামগুলো বাংলায় ইউনিকোডকে সমর্থন করে না। ডিজিটাল যন্ত্রে বাংলার ব্যবহারের ক্ষেত্রে এটাও একটি বড় সংকট।

ইউনিকোড এবং ইউনিকোড কনসোর্টিয়াম কী

ইউনিকোড হলো সারা দুনিয়ার ভাষাসমূহের জন্য ডিজিটাল যন্ত্রের এক ও অভিন্ন এনকোডিং ব্যবস্থা। অন্যদিকে ইউনিকোড কনসোর্টিয়াম হলো এই কোড প্রমিত করার জন্য মার্কিন যুক্তরাষ্ট্রে অবস্থিত একটি অলাভজনক সংস্থা। এর সূচনা ১৯৮৬ সালে হলেও এর জন্ম ১৯৮৭ সালে। ১৯৮৮ সালে এটি মোটামুটি একটি রূপ পরিগ্রহ করে। ১৯৯১ সালের জানুয়ারিতে ইউনিকোড ইনকর্পোরেটেড হয়। রয়াল জেরস্কেস জো বেকার এবং অ্যাপল কমপিউটারের লি কলিন্স ও মার্ক ডেভিস প্রাথমিক আলোচনার মধ্য দিয়ে ইউনিকোড কনসোর্টিয়ামের যাত্রা শুরু ▶

করেন। মূলত র‍্যাঙ্ক জেরক্স ও অ্যাপল কমপিউটারের উদ্যোগ হলেও এটি এখন সারা দুনিয়ার ভাষাসমূহের ডিজিটাল যন্ত্রের মান নির্ণয় করে থাকে। যদিও আইএসও (ইন্টারন্যাশনাল স্ট্যান্ডার্ড অর্গানাইজেশন) হলো বিশ্বের মান নির্ধারণ বা প্রমিতকরণ সংস্থা, তথাপি ইউনিকোড কনসোর্টিয়ামই কার্যত কমপিউটারে বিশ্বের ভাষাসমূহ ব্যবহার করার একমাত্র মান প্রস্তুতকারী সংস্থা। এখনকার প্রেক্ষিতে এই দুটি সংস্থাকে আলাদা করে ভাবার কোনো অবকাশ নেই।

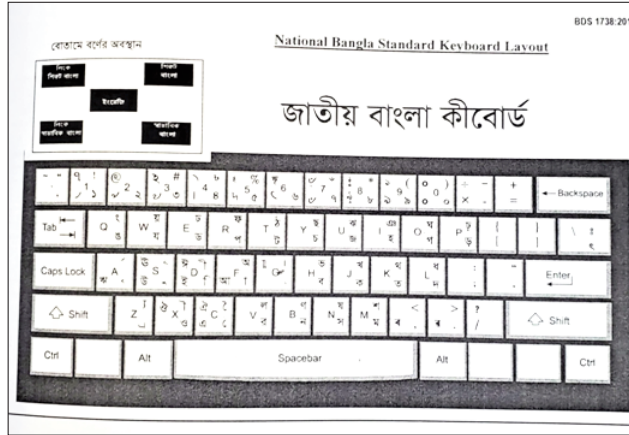
ইতিহাসের পাতায় চোখ রাখলে জানা যাবে, ইউনিকোড শব্দটির উৎপত্তি জো বেকারের হাতে। তিনি from unique, universal, and uniform character encoding-এর বিবেচনায় এই নামটি রাখেন। বস্তুত এর প্রথম মানটি তৈরি করার কাজ শুরু হয় ১৯৮৬ সালে। কিন্তু সেই সংস্করণটি প্রস্তুত করতে সময় লাগে ছয় বছর। তাই ১৯৯২ সালের জুন মাসে প্রথম সংস্করণটি প্রকাশিত হয়। এর মাঝে ইউনিকোডের ১২টি সংস্করণ প্রকাশিত হয়েছে। ২০১৯ সালের মে মাসে প্রকাশিত হয় এর ১২.১ সংস্করণ।

ইউনিকোড ওয়েবসাইট থেকে প্রাপ্ত তথ্য অনুসারে এ ইউনিকোড ১২.১ সংস্করণের বিবরণ এ রকম- মোট ১,৩৭,৯২৯টি বর্ণ রয়েছে।

কমপিউটারের যারা আদি ব্যবহারকারী, তারা জানেন এই ডিজিটাল যন্ত্রটি একেবারে শুরুতে

রোমান হরফ ছাড়া আর কোনো হরফ বুঝত না। শুরুতে আমরা এসব প্রসঙ্গ এবং আসকি এনকোডিং নিয়ে কথা বলেছি। র‍্যাঙ্ক জেরক্স কোম্পানি আসকির কোডসীমা অতিক্রম করার চেষ্টা করলেও একই সমান কোড বহাল রেখে রোমান ভাষার পাশাপাশি অন্য ভাষা ব্যবহারের সহজ সুযোগ আসে ১৯৮৪ সালে জন্ম নেয়া মেকিন্টোস কমপিউটারে। ১৯৮১ সালে জন্ম নেয়া আইবিএম পিসির জন্য প্রণীত ডস অপারেটিং সিস্টেমেও রোমান হরফের কোড সীমানায় অন্য ফন্ট এবং সেই সুবাদে অন্য ভাষা ব্যবহার করা যেত। কিন্তু প্রক্রিয়াটি জটিল ছিল বলে অ্যাপল কমপিউটার কোম্পানির মেকিন্টোস কমপিউটার ডেস্কটপ বিপ্লবে অগ্রনায়কের ভূমিকা পালন করে। অ্যাপলের গ্রাফিক্যাল ইউজার ইন্টারফেস এবং কুইক ড্র অ্যাপ্লিকেশনগুলো খুব সহজে গ্রাফিক পদ্ধতিতে রোমান ছাড়া অন্য ভাষার বর্ণমালা উৎপন্ন করতে শুরু করে। মাল্টিপল ফন্ট ব্যবহারের প্রযুক্তিও এজন্য সহায়তা করেছে। ফলে ডস অপারেটিং সিস্টেমে বাংলা প্রয়োগ করার চেষ্টা আশির দশকের শুরুতে করা হলেও বস্তুত অ্যাপল মেকিন্টোস কমপিউটারেই প্রথম বাংলা ভাষা সঠিক ও সুন্দরভাবে লেখা হয়। ১৯৮৭ সালের ১৬ মে প্রকাশিত হয় কমপিউটারে কম্পোজ করা প্রথম বাংলা পত্রিকা- আনন্দপত্র। পরবর্তী সময়ে ১৯৯৩ সালে উইডোজের হাত ধরে মেকিন্টোসের সেই প্রযুক্তি পিসিতে আসে। শুধু ইন্টারনেট ছাড়া অন্য সব কাজেই আমরা এখনও সেই প্রযুক্তিই ব্যবহার করছি।

২৫৬ কোডের আসকি কোডসেট নিয়ে বাংলা লেখার প্রধান অসুবিধা তিনটি। প্রথমত বাংলা বর্ণের সংখ্যা অনেক বেশি। ২৫৬ কোডের মাঝে সর্বোচ্চ ২২০টি কোড ব্যবহার করে বাংলা সব যুক্তাক্ষর অবিকৃতভাবে লেখা অসম্ভব। এতে বর্ণসমূহের সৌন্দর্য রক্ষা করা কঠিন হয়ে পড়ে। বেশিরভাগ বর্ণকে সঠিকভাবে উৎপন্ন করা গেলেও অনেক বর্ণ তার প্রকৃতরূপে দেখা যায় না। বিশেষ করে যুক্তাক্ষরের ক্ষেত্রে সমস্যাটি গভীর। ফলে যখনই আমরা বাংলা হরফের বিষয়টি পর্যালোচনা করি, তখনই দেখি যুক্তাক্ষরগুলো তার আদি রূপ হারিয়ে ফেলেছে বা নানা ধরনের অক্ষরাংশ দিয়ে তৈরি হচ্ছে বলে যথাযথভাবে সেটি উৎপন্ন হচ্ছে না। বাংলা হরফে যেহেতু শুধু বর্ণ নয় ফলা এবং চিহ্নও আছে, সেহেতু সেগুলো সঠিকভাবে



বর্ণসমূহের সাথে যুক্ত হচ্ছে কিনা সেটিও দেখার বিষয়। বাস্তবতা হলো, ২৫৬টি কোডে বাংলা সঠিকভাবে লেখা সম্ভব নয়। আমরা যারা বাংলা হরফ তৈরি করি তারা কার্যত আপস করি। একে মন্দের ভালো বলা যায়।

আসকি কোডের দ্বিতীয় অসুবিধা ছিল যে বাংলা ভাষার জন্য কোনো আসকি মান ছিল না। বাংলাদেশের বিএসটিআই এক সময়ে বাংলার জন্য একটি মান তৈরির চেষ্টা করে ব্যর্থ হয়। যদিও ২০০০ সালে একটি মান তৈরি হয় (বিডিএস ১৫২০:২০০০) কিন্তু সেটি আইএসও গ্রহণ করেনি। অন্যদিকে বেসরকারি সফটওয়্যার নির্মাতারা নিজেরা একেকটি মান তৈরি করে বলে একাধিক মানের জন্ম হয়েছে। এজন্য বাংলাদেশের একাধিক মান, পশ্চিমবঙ্গের একাধিক মান এবং আসাম-ত্রিপুরার মানগুলো নিয়ে আমরা ব্যাপকভাবে সংকটে পড়েছি। এতদিনে '১৮ সালে বাংলাদেশ আসকির জন্যও মান নির্ধারণ করেছে। সুতরাইএমজে ফন্টের কোড এখন বিডিএস ১৯৩৫:২০১৮ মানে প্রমিত হয়েছে।

খুব বেশি গুরুত্বপূর্ণ না হলেও তৃতীয় আরেকটি সমস্যা হলো- আসকির একই কোড ইংরেজি ও বাংলা ব্যবহার করে বলে ফন্ট পরিবর্তনের সাথে সাথে ভাষাও বদলে যেতে পারে। বাংলা ও ইংরেজি মিশ্রিত দলিলে এটি বেশ বড় ধরনের সমস্যা তৈরি করে। ফন্ট বদলাতে গেলে বা কনভার্ট করতে গেলে সেগুলো ভীষণ সমস্যা তৈরি করে। আগামীতে এ বিষয়ে আরও আলোচনা করব।

রোমান হরফে লেখা হয় এমন ভাষা ছাড়া

দুনিয়ার সব ভাষাই আসকি কোড নিয়ে যন্ত্রণায় ভুগেছে। ভারতীয় ভাষাগুলো ছাড়াও জাপানি, কোরীয় বা চীনা ভাষার যন্ত্রণা আরও অনেক বেশি ছিল। তবে সবার মতো ইউনিকোড মান আমাদের আসকি এনকোডিং সমস্যার সমাধান দিয়েছে। যদিও ইউনিকোড মানে শুধুমাত্র মূল বর্ণগুলোর মান রয়েছে এবং আমাদের যুক্তবর্ণকে কোনো মান দেয়া হয়নি, তথাপি ওপেনটাইপ নামক একটি প্রযুক্তি দিয়ে আমরা অক্ষরের সংখ্যা যেকোনো পর্যায়ে নিয়ে যেতে পারি। ফলে অক্ষরের আকার আকৃতি নিয়ে যে সমস্যা আসকিতে রয়েছে সেটি আর থাকে না। ইউনিকোড যেহেতু সবার কাছে গ্রহণযোগ্য সেহেতু এর মান নিয়ে কোনো প্রশ্নও উঠেছে না। বাংলাদেশে ইউনিকোড মান সরকারি স্বীকৃতি পেয়েছে। এমনকি পশ্চিমবঙ্গেও এটি স্বীকৃতি পেয়েছে। পশ্চিমবঙ্গের যে সংস্থাটি ইউনিকোডের সদস্য তাদের মন্তব্য হলো- Data Representation and Encoding Standard: Society for Natural Language Technology Research has decided to accept UNICODE 5.0 and upwards as the standard for data representation and encoding for Bangla.

এর মানে হচ্ছে, ইউনিকোড মান ব্যবহার করলে বাংলাদেশের একটি টেক্সট ভারতের পশ্চিমবঙ্গেও একইভাবে কোনো পরিবর্তন ছাড়া ব্যবহার করা যাবে। অন্যদিকে ইউনিকোডের মানে

যেহেতু ভাষাগুলোর কোডিং আলাদা সেহেতু কোনো টেক্সটের ফন্ট বদলালে সেটির ভাষাও বদলে যায় না। তবে অত্যন্ত বিনয়ের সাথে আমি একথা বলব, আমাদের ইউনিকোডের সদস্যপদ অনেক আগেই গ্রহণ করা জরুরি ছিল। এর প্রথম কারণ হচ্ছে বাংলা আমাদের রাষ্ট্রভাষা। ফলে এই ভাষার মান নির্ণয়ের ক্ষেত্রে আমরা নীরব থাকতে পারি না। আমি স্মরণ করতে পারি, শুরু থেকেই ইউনিকোড মান যখন নির্ধারিত হতে থাকে তখন বাংলাদেশের প্রতিনিধিত্ব ছিল না বলে তাতে আমরা যেভাবে বাংলা লিখি এবং আমাদের যেসব হরফের খুবই প্রয়োজন সেগুলো ইউনিকোড অন্তর্ভুক্ত করেনি। প্রধানত হিন্দির অনুকরণে বাংলার প্রমিতকরণ করা হয়। ফলে প্রথম দিকের সংস্করণগুলোতে আমাদের ড, ঢ, য় ও ঙ ছিল না। ইউনিকোডে এসব অক্ষর প্রবেশ করতে অনেক সময় লেগেছে। আমি স্মরণ করতে পারি, থাইল্যান্ডে যখন আমি ইউনিকোডের প্রথম সংস্করণের নমুনা দেখি তখন স্পষ্টতই বাংলাকে হিন্দির অনুরূপ মনে হয়। সেখানে প্রতিবাদ করার ফলে প্রথম সংস্করণ চূড়ান্ত করার আগেই অনেক অগ্রগতি হয়।

আমাদের পক্ষ থেকে সরকারকে আমরা বারবার এই কথা বলে এসেছি, ইউনিকোড কনসোর্টিয়ামের সদস্যপদ পাওয়াটাই বড় কথা নয়। এতে সক্রিয়ভাবে অংশ নেয়াও খুব প্রয়োজনীয়। ২০১০ সালের সদস্যপদ লাভের পর থেকে এই কনসোর্টিয়ামের নানা ধরনের কর্মকাণ্ড হয়েছে। কিন্তু এই সংস্থায় আমাদের যতটা সক্রিয় হওয়া উচিত ছিল ততটা আমরা হইনি।

বাংলা কোড ও ইউনিকোড কনসোর্টিয়াম

বাংলাদেশ রাষ্ট্রের ভিত্তি হিসেবে আমরা বাংলা ভাষাকেই গণ্য করি। বস্তুত ১৯৪৭ সালে ধর্মভিত্তিক মুসলিম জাতীয়তার নামে পাকিস্তান তৈরির পর ১৯৪৮ সালেই বাঙালিরা ভাষার প্রশ্নে জেগে ওঠে। ১৯৫২ সালে বাংলাকে পাকিস্তানের অন্যতম রাষ্ট্রভাষা হিসেবে প্রতিষ্ঠা করার আন্দোলনের সফলতা থেকেই এই জাতি তার জাতির রাষ্ট্র বাংলাদেশ প্রতিষ্ঠা করার আন্দোলনের ভিত রচনা করে। কালক্রমে সেই জাতির রাষ্ট্র প্রতিষ্ঠিত হয় এবং বাংলা পাকিস্তানের অন্যতম রাষ্ট্রভাষা থেকে বাংলাদেশ রাষ্ট্রের একমাত্র রাষ্ট্রভাষায় পরিণত হয়। সেই রাষ্ট্রভাষা বাংলার ডিজিটাল যন্ত্রে প্রমিতকরণের জন্য সবচেয়ে প্রয়োজনীয় উপাদানটি ছিল অলাভজনক প্রতিষ্ঠান ইউনিকোড কনসোর্টিয়ামের সাথে যুক্ত হওয়া। এই সংস্থাটি রোমানসহ দুনিয়ার সব ভাষার সব হরফের প্রমিতকরণ করে থাকে। বাংলা ভাষার জন্য ব্যবহৃত লিপিমালার প্রমিতকরণও এই সংস্থাটি করেছে। কিন্তু যথাসময়ে সেই সংস্থার সাথে আমরা যুক্ত হতে পারিনি বলে বাংলা বর্ণমালার প্রমিতকরণ নিয়ে সংকট ছিল। এই প্রতিষ্ঠানের সাথে যুক্ত হওয়ার প্রধান ও একমাত্র উপায়টি ছিল সরকারি কোনো প্রতিষ্ঠান কর্তৃক এর সদস্য হওয়া। সদস্য হওয়ার জন্য প্রয়োজন ছিল মাত্র ১২ হাজার মার্কিন ডলার বা প্রায় সাড়ে ৯ লাখ টাকা। কিন্তু অতীতের বাংলাদেশ সরকার বছরের পর বছর ধরে এই ১২ হাজার ডলার দিয়ে ইউনিকোড কনসোর্টিয়ামের সদস্যপদ গ্রহণ করেনি। সাধারণ মানুষের পক্ষ থেকে সরকারের কাছে বছরের পর বছর আবেদন নিবেদন করা হয়েছে। কিন্তু এই শতকের প্রথম দিকে ক্ষমতাসীন চারদলীয় জোট সরকার বা তার পরের তত্ত্বাবধায়ক সরকার কেউ বাংলা ভাষার জন্য এই সামান্য কাজটুকু করতে চায়নি। আমরা কোনোভাবেই সেইসব সরকারকে এই সদস্যপদের গুরুত্ব বোঝাতে পারিনি। তারা এই প্রতিষ্ঠানটি কী, সেটিই বুঝেনি।

শেষ পর্যন্ত জাতির জনক বঙ্গবন্ধুকন্যা শেখ হাসিনার সরকার এবং স্থপতি ইয়াফেস ওসমানের মন্ত্রিত্বের সময়ে সেই স্বপ্নের সিঁড়ি আমরা পার হলাম। ইউনিকোডের জন্মের প্রায় ২৩ বছর পর বাংলাদেশ সরকার ইউনিকোড কনসোর্টিয়ামের সদস্য হলো। অথচ অন্তত আরও ১৯ বছর আগে এই সদস্যপদ গ্রহণ করা যেত। বিশেষ করে ২০০১ সালের পরে ইউনিকোড কনসোর্টিয়ামের সদস্যপদ লাভ করা আমাদের জন্য জীবন-মরণ ব্যাপার ছিল। এটি না হওয়ার ফলে বাংলা ভাষার অগ্রগতি আর যাই হোক না কেন, বহু বছর পিছিয়ে গেছে— এ বিষয়ে সন্দেহ করার কিছু নেই।

অবশেষে ৩০ জুন ২০১০ ইউনিকোড কনসোর্টিয়াম বাংলাদেশকে সদস্যপদ দান করে এবং ১ জুলাই থেকে এই সদস্যপদ সক্রিয় হয়। প্রাথমিকভাবে এক বছরের জন্য বাংলাদেশ ১২ হাজার ডলার সদস্য চাঁদা প্রদান করে এবং প্রতি বছর বার্ষিক চাঁদা পরিশোধ করে এই সদস্যপদ নবায়ন করা সম্ভব হচ্ছে। ১৮ মার্চ ২০১০ ইউনিকোড সদস্যপদের জন্য আবেদন করে ৩০ জুনের মাঝে সদস্যপদ প্রাপ্তি নিঃসন্দেহে একটি

বড় ধরনের মাইলফলক অর্জন। এর ফলে প্রমাণিত হলো যে, আন্তরিকতা থাকলে কোনো কাজই অসম্ভব নয়। ব্যক্তিগতভাবে ধন্যবাদ দেয়া উচিত বিসিসি তৎকালীন নির্বাহী পরিচালক মাহফুজুর রহমানকে।

যাই হোক, ইউনিকোড কনসোর্টিয়ামের সদস্যপদ লাভের পর কমপিউটার কাউন্সিলকে সক্রিয় করে আমরা আমাদের সমস্যাগুলো তুলে ধরার চেষ্টা করি। কিন্তু ইউনিকোড এনকোডিং নিয়ে আমাদের যে সমস্যা ছিল সেই সমস্যার সমাধান এখনও হয়নি। ২০১৪ সালের মে মাসে কমপিউটার কাউন্সিলর এনামুল কবির ইউনিকোডের সভায় প্রথমবারের মতো যোগ দেন এবং বাংলা নিয়ে বাংলাদেশের সমস্যার বিষয়টি উপস্থাপন করেন। এরপর তিনি আরও একবার একটি সভায় যোগ দেন। কিন্তু '১৭ সাল অবধি বিসিসি প্রায় নিশ্চুপই থেকেছে।

ইউনিকোড নিয়ে বাংলাদেশের সমস্যাটি খুব বড় নয়। এর প্রধানতম কারণ হলো ইউনিকোড কনসোর্টিয়ামে আমাদের অনুপস্থিতি। ভারত থেকে যেহেতু ভারতীয় ভাষাসমূহের জন্য কোড প্রস্তাব করা হয় সেহেতু প্রথমে আমাদের কিছু প্রয়োজনীয় বর্ণ যুক্ত হয়নি। অন্যদিকে আমাদের দাঁড়ি ও দুই দাঁড়িকে দেবনাগরীর কোড থেকে নেয়ার জন্য প্রমিত করা হয়। খুব সঙ্গত কারণেই এই বিষয়ে ভারতের কোনো আপত্তি থাকার কথা নয়। কিন্তু আমাদের আপত্তি আছে। দেবনাগরী ও বাংলাসহ ভারতীয় ভাষাসমূহের এক ধরনের হরফ থাকার পরেও আমরা আলাদা আলাদা কোড যদি নিতে পারি তবে দাঁড়ি আর দুই দাঁড়ির কোড কেন বাংলায় আলাদা হতে পারবে না?

এই প্রেক্ষিতেই আমরা ২০১১ সালে আমাদের নিজস্ব মানে এই কোড দুটি অন্তর্ভুক্ত করি (বিডিএস ১৫২০:২০১১) এবং ২০১৪ সালের মে মাসে আমাদের প্রতিনিধি ইউনিকোড কনসোর্টিয়ামকে আমাদের প্রস্তাব মেনে নেয়ার অনুরোধ করেন। সেই সময়ে ইউনিকোড কনসোর্টিয়াম তাতে সম্মত না হলেও পরে মেইল পাঠিয়ে জানায় যে, ইউনিকোডের ৮ম সংস্করণে তারা আমাদের প্রস্তাব মূল্যায়ন করবে। গত ৯ জুলাই ২০১৪ ইউনিকোড কনসোর্টিয়াম বাংলাদেশ কমপিউটার কাউন্সিলের মোহাম্মদ এনামুল কবিরের কাছে একটি মেইল পাঠিয়েছে। কবির সজ্জন মানুষ বলেই মেইলটির একটি কপি আমাকে ফরোয়ার্ড করেছেন। এটি রাষ্ট্রীয় গোপন তথ্য নয় এবং বাংলাদেশের বাংলা ভাষাভাষী মানুষের জন্য একটি অর্জন বলে আমি এই মেইলটি প্রকাশ্যে শেয়ার করছি।

এটি আমাদের জন্য, বাংলা ভাষার জন্য, বাংলাদেশ সরকারের জন্য একটি বিশাল অর্জন। মেইলটি এ রকম— Hello again Mohammad,

We have taken editorial action items in the Editorial Committee for this. The editors will update the annotations in the names list for Version 8. (Version 7 has already been published.)

We will also update the Bengali block description for the Version 7 Core Specification to note the names dahri and double dahri used in Bangla. (That would

be in the section corresponding to Chapter 9, Section 2.) The Core Spec is planned for October 2014 publication.

If you think it's necessary to have any further information in meeting minutes beyond what is now documented, then we could enter your recent comments as feedback on 7.0, and capture formal action items in the August meeting. (That is coming up in about 1 month.) I hope that helps.

Regards, Rick

আমরা ইউনিকোড কনসোর্টিয়ামের ৮.০ সংস্করণ দেখেছি যেখানে দেবনাগরীর দাঁড়ি ও দুই দাঁড়িকে বাংলারও কোড বলে বিবরণ লেখা হয়েছে। These punctuation marks are for common use for the scripts of India despite being named “DEVANAGARI”. They also occur as abbreviation signs in some South Indian scripts. বাংলার কোডচার্টে লেখা আছে এভাবে— For viram punctuation, use the generic Indic 0964 and 0965. Note that these punctuation marks are referred to as dahri and double dahri in Bangla. 09E4 “ → 0964 | devanagari danda 09E5 “ → 0965 || devanagari double danda.

তবে এতে আমাদের দাবি পূরণ হয়নি। আমরা এই দুটি বর্ণকে বাংলার আলাদা কোড হিসেবে পেতে চাই। দুর্ভাগ্যজনকভাবে ইউনিকোড কনসোর্টিয়াম আমাদের জন্য রিজার্ভ করা কোড দুটিকে স্বীকৃতি দিচ্ছে না। অথচ দেবনাগরীর জন্য এন্সটেনেডেড কোড বরাদ্দ করা হয়েছে। দুর্ভাগ্যজনকভাবে আমাদের সচেতনতার অভাবে আমরা ইউনিকোড মানতে গিয়ে নিজেদের মানই প্রয়োগ করিনি। সরকারের প্রায় সব মন্ত্রণালয় ও বিভাগ ইউনিকোড মান মেনে চলতে চাইলেও বাংলাদেশের মান মেনে চলে না। এর বাইরেও ইউনিকোড প্রচলন নিয়ে সরকারের সংকট আছে নানামুখী।

আমাদের নিজেদের জন্য একটি সস্তষ্টির বিষয় হলো '১৭-১৮ সালে আমরা বাংলার তিনটি প্রমিত মান তৈরি করতে পেরেছি। বাংলা আসকি কোডের প্রমিত মান বিডিএস ১৯৩৫:২০১৮ (সুতস্বীএমজে ফন্টের কোড), বাংলা কিবোর্ডের মান বিডিএস ১৭৩৮:২০১৮ (এটি বিজয় কিবোর্ড ৩.০) এবং বিডিএস ১৫২০:২০১৮ ইউনিকোড ১২.০-এর সংশোধিত মান। যদিও সরকারি অফিসগুলো বিডিএস ১৫২০:২০১৮ ছাড়া বাকি দুটি মান এখনও মানছে না, তবুও প্রমিত করার ফলে নিশ্চয়ই এসব মান সর্বজনীন মান হিসেবে ব্যবহার হবে।

ইউনিকোড কর্তৃপক্ষ দাঁড়ি ও দুই দাঁড়ির জন্য ইউনিকোডে রিজার্ভ কোড রেখেছে। অন্যদিকে ড়, ঢ়, য়-এর জন্যও রিজার্ভ কোড রেখেছে। কিন্তু ভারতের একটি মহল এসব রিজার্ভ কোড কোনোটিকেই ইন্টারনেটে ডোমেইনে লিখতে দিচ্ছে না এবং নিউ ব্রাহ্মী লিপি প্রক্রিয়ায় এর বিরোধিতা করছে। আমাদেরকে যেকোনো মূল্যে আমাদের অধিকার আদায় করতেই হবে এবং বাংলা ভাষাকে অবিকৃত রাখতে হবে

ফিডব্যাক : mustafajabbar@gmail.com