সার্বিক সমস্যা সমাধানসহ

কমপিউটারে বাংলা ব্যবহার

মোঃ মারুফ হাসান
শেষ বর্ষ, কমপিউটার প্রকৌশল
ছিংহুয়া বিশ্ববিদ্যালয়
সহযোগী ঃ মোঃ কাওসার, আশরাফুল হক

কমপিউটার জগৎ-এ কমপিউটারে বাংলা ব্যবহার সংক্রান্ত একাধিক লেখা ছাপানো হয়েছে, তাই আমার লেখাকে সংক্ষিপ্ত রাখার জন্য আমি ভূমিকাকে সংক্ষিপ্ত করলাম। আগ্রহী সকলেই পূর্ব প্রকাশিত লেখা থেকে মোটামুটি বাংলা ভাষা কমপিউটারাইজ সংক্রান্ত সমস্যাসমূহ সম্পর্কে ইতিমধ্যে ওয়াকিবহাল।

আমাদের কমপিউটারে বাংলা ব্যবহারের পদ্ধতিটির সাথে অন্য পদ্ধতিগুলোর মূল তফাৎ হল আমরা Double Byte Character System (DBCS) ব্যবহার করছি।

আমাদের কথাটি লেখার কারণ হল আমি ও আমার তিনজন স্বদেশী ছাত্রবন্ধু এবং এক জন চীনা অধ্যাপকের সমন্বয়ে পাঁচজনের একটি গ্রুপ গত ১লা জুন ৯২ থেকে প্রজেক্টটি শুরু করেছি। কমিটিটি নিম্নরূপ ঃ

Double Byte Character System— Bengali in Computer Task Groups (সংক্ষেপে BCTG)

Chief Advisor : Prof. Mao Yu Hang.
Project Manager : Md. Maruf Hasan.
Algorithm & Programming : Maruf.
Linguistic Analysts: Kawsar, Ripon, Anil Krishna.
Graphic Library Design : Ripon, Maruf.
Manual writing & Public Relation : Kawsar, Maruf.
Database Library Creation : Anil, Ripon, Kawsar, Maruf.

উদ্দেশ্য ঃ

১। কমপিউটারে বাংলা ব্যবহারকে ওয়ার্ডপ্রসেসিং এর গণ্ডি থেকে বের করতে হবে। অর্থাৎ Database Management, Data Communication (Network), Text to Speech (কমপিউটার যে কোন ধরনের ডকুমেন্ট পড়তে পারবে), Bengali computer for Blind people (অন্ধ মানুষের জন্য কমপিউটার বিশেষ প্রক্রিয়ায় বাংলা প্রিন্ট করবে যা অনুভূতি ইন্দ্রিয়ের দ্বারা একজন অন্ধ মানুষ বুঝতে (পড়তে) পারবে।

২। শুধু বাংলা ও ইংরেজী নয় যে কোন ভাষায় (কমপক্ষে দশটি ভাষার) কম্পোজ ও ম্যানেজমেন্টের সুবন্দোবস্ত রাখা হবে আমাদের সিস্টেমে। এতে করে বিভিন্ন ভাষায় ডাটা এন্ট্রির কাজ করা যাবে। এছাড়া পররাষ্ট্র দপ্তর, দূতাবাস ও বহুজাতিক কোম্পানীসমূহের জন্য এ ধরণের সিস্টেমের গুরুত্ব অপরিসীম।

৩। আমাদের দেশীয় প্রযুক্তিতে হার্ডওয়্যার তৈরীর আশা ক্ষীণ তাই কী বোর্ড হোক আর প্রিন্টারই হোক কোনটিতেই কোনক্রমে নতুন হার্ডওয়্যার ব্যবহারের চিন্তা করা ঠিক হবে না। এতে দেশী অর্থের অপচয় হবে। সেক্ষেত্রে সফটওয়্যারে সমাধা হলে দেশের অর্থনীতির জন্য অনুকূল হবে।

৪। কমপিউটারে বাংলা ব্যবহারের পদ্ধতিটি হবে User Transparent অর্থাৎ জটিলতাসমূহকে টাইপিষ্ট বা অপারেটরের দৃষ্টির বাইরে রাখতে হবে। একটি পরিচ্ছন্ন কীবোর্ডের সামনে বসে আয়াসে তাঁরা কাজ করবেন।

৫। অক্ষর বা যুক্তাক্ষরের আকৃতি পরিবর্তনের মাধ্যমে বাংলা ভাষার ঐতিহ্য খণ্ডন (প্রযুক্তিগত বাধা থাকলে) অযৌক্তিক নয়। তবে একেবারে নাকচ না হয়ে এ সিদ্ধান্ত নেওয়া উচিত হবে না।

৬। বর্তমানের যান্ত্রিক টাইপ রাইটার-এর কীবোর্ড লে-আউটকে অক্ষুন্ন রাখতে চাইলে সে ব্যবস্থা করা কষ্টকর নয়। তবে যেহেতু এক দশকের মধ্যেই যান্ত্রিক টাইপ রাইটারের ব্যবহার উল্লেখযোগ্যভাবে কমে যেতে বাধ্য, তাই সেখানে যুক্তিগ্রাহ্য পরিবর্তন ঘটলে ক্ষতি নেই এবং পরবর্তীতে এই পরিবর্তিত লেআউট-এ ইলেকট্রনিক টাইপ রাইটারের লেআউট নির্ধারণ করা যুক্তিসঙ্গত।

৭। শুধুমাত্র একবার কী চেপে কমপিউটারকে বাংলা থেকে ইংরেজী এবং ইংরেজী থেকে বাংলা মোডের পরিবর্তন সম্ভব করতে হবে।

৮। High (Bengali Character) Input rate পাবার ব্যবস্থা থাকতে হবে অর্থাৎ অল্প সংখ্যক কী চেপে বেশী ইনপুট পাবার দিকে লক্ষ্য রাখতে হবে।

৯। পাবলিশিং-এর জন্য গাণিতিক চিহ্নসহ অন্যান্য সমস্ত চিহ্ন ও সাবস্ক্রীপ্ট, সুপারস্ক্রীপ্ট সুবিধা থাকতে হবে। সহজে ছক বা টেবিল তৈরীর সুবন্দোবস্ত থাকতে হবে।

১০। স্বল্প ব্যয়ে Hardware Implementation যাতে সম্ভব হয় তার প্রতি কঠোর দৃষ্টি রাখতে হবে।

১১। যে কোন ইংরেজী সফটওয়্যার স্বল্প পরিবর্তনের মাধ্যমে বাংলা ব্যবহার করার কথা মনে রাখতে হবে।

১২। আইবিএম পিসি এক্সটি, এটি, ২৮৬, ৩৮৬ ইত্যাদি কমপিউটারে, মনোক্রম হারকিউলিস গ্রাফিক্স কার্ডে অথবা সিজিএ, ইজিএ, ভিজিএ ইত্যাদি ডিসপ্লে ডিভাইসে; নয় পিন থেকে ২৪ পিন যে কোন ডট মেট্রিক্স প্রিন্টারে ব্যবহার উপযোগী হবে। অর্থাৎ সিস্টেমটি পোর্টেবল হবে।

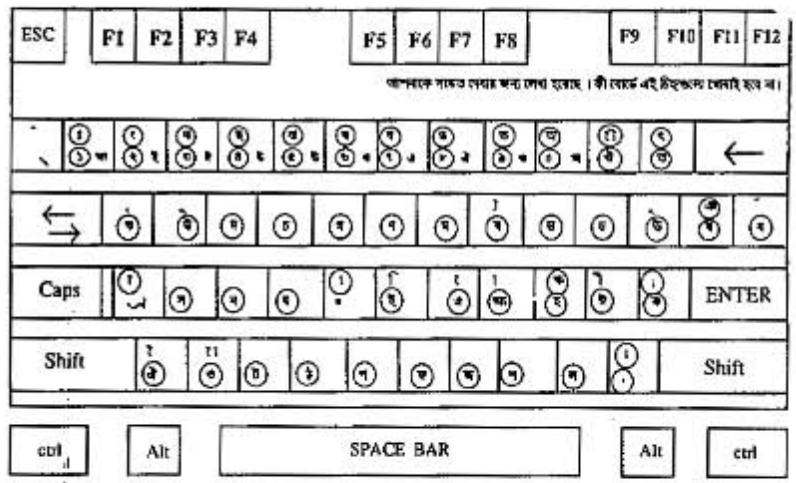
১৩। একাধিক ফন্ট, ষ্টাইল ও সাইজে বাংলা হরফ প্রিন্টিং-এর সুযোগ থাকতে হবে।

মাসাধিককালের অবিরাম প্রচেষ্টা ও গ্রুপের প্রত্যেকটি সদস্যের কর্তব্যনিষ্ঠার জন্য আমাদের উদ্দেশ্য সফল হতে চলেছে। আমাদের কীবোর্ড লেআউট নিম্নরূপ। তবে যে কোন ধরণের কী বোর্ড লে-আউট কয়েক ঘন্টার মধ্যে তৈরী করার মত হাতিয়ার আমাদের ইতিমধ্যে তৈরী হয়েছে। আমাদের সিস্টেমের supplement হিসেবে এই হাতিয়ারকে যুক্ত করলে ব্যবহারকারী স্বাধীনভাবে তার কীবোর্ড লেআউট তৈরী করতে পারবে।

ইনপুট পদ্ধতি ঃ

(চিত্র – ১ এর কীবোর্ড দ্রষ্টব্য)

আপনার কমপিউটার চালু করার পর আমাদের সফটওয়্যার রান করুন। Alt-F* (F* বলতে যে কোন একটি ফাংশন কী বুঝানো হলো) চাপ দিন। কী বোর্ডে বৃত্তের মধ্যে দেখানো প্রতিটি অক্ষরকে নির্দিষ্ট কী (অথবা Shift এর সাথে কী) এবং স্পেসবার চাপলেই পাবেন। ড়, ঢ়, য়, ঢ় এর জন্য যথাক্রমে ব-০, ড-০, য-০, ঢ-০ ইত্যাদি চাপতে হবে; এটি একটি ব্যতিক্রম। এছাড়া আ-কার; ই-কার, সংযুক্ত অক্ষর কিভাবে পাওয়া যাবে তার stroke order নিম্নের ছকে কিছু নমুনা দেখানো হল (চিত্র নং-২ দ্রষ্টব্য)।



[চিত্র নং ঃ ১]

দ্রষ্টব্য ঃ

১। যুক্তাক্ষর বলুন আর বিকৃত উ-কার (যেমন রু) র-ফলাকার (ব্রা) ইত্যাদি ব্যতিক্রমসমূহ নির্ঝঞ্ঝাটভাবেই ডিসপ্লে ও ডট ম্যাট্রিক্স প্রিন্টারে অবিকল হাতে লেখার মতই ফুটে উঠবে।

২। ি ী ইত্যাদিকে ঠিক অক্ষরের মাথার উপরেই পাবেন। ু ূ ইত্যাদিকে ঠিক অক্ষরের নীচে পাবেন।

৩। লক্ষণীয় যে ক্র এর typical রূপ অপরিবর্তিতই থাকবে, তদ্রূপ আপনি থ-কে দুবার চেপে থ-হুবহু ডিসপ্লে ও প্রিন্ট করতে পারবেন। গ ও ট চেপে ঠিক গু-ই পাবেন; গ্ নয়।

কোডিং পদ্ধতি ঃ

(চিত্র নং-৪ এর নিম্নাংশ দ্রষ্টব্য)

বাংলা ভাষার ৩৯টি ব্যঞ্জনবর্ণ ও ১১টি স্বরবর্ণ আছে। এরা বিভিন্ন রকম বিন্যাসের মাধ্যমে (যেমন ব+আ = বা, ক + ষ-ফলা = ক্ষ)

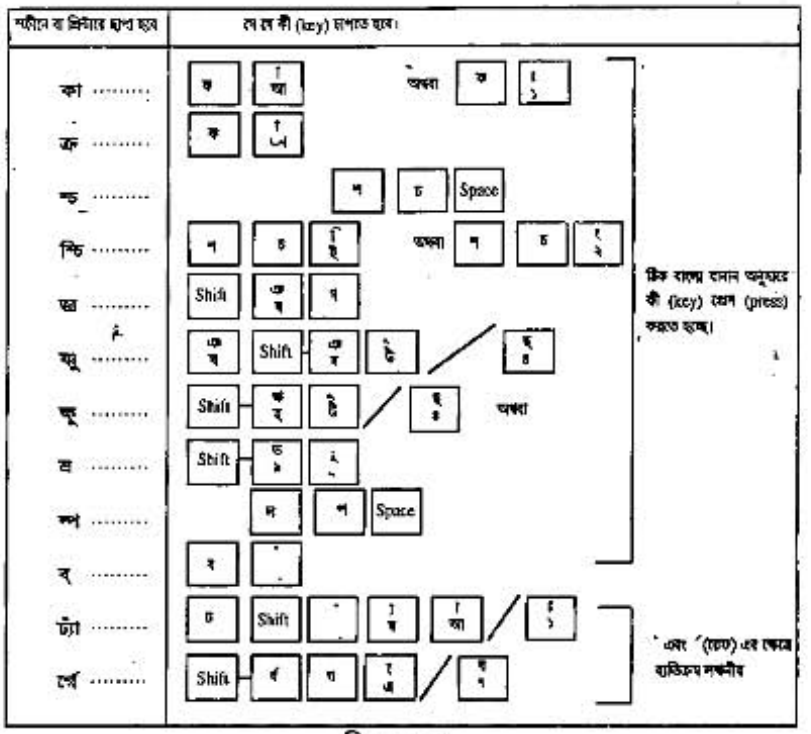
বিভিন্ন সাউন্ড তৈরী করে। এছাড়া এদের কম্বিনেশনের মাধ্যমে যুক্তাক্ষর তৈরী হয়। যুক্তাক্ষর ও আবার বিভিন্ন স্বরাকারের সমন্বয়ে নতুন শব্দ গঠন করে। এছাড়া হসন্ত (্) চন্দ্রবিন্দু (ঁ) ইত্যাদি ও ব্যঞ্জনবর্ণের সাথে যুক্ত হয়ে নতুন শব্দ তৈরী করে।

আনুমানিক বাংলা ভাষাতে এ ধরনের ৩০ × ১৩ = ৩০০ ব্যঞ্জনবর্ণ-স্বরবর্ণ, ৩৫০ × ৫ = ১৭৫০ যুক্তবর্ণ - স্বরবর্ণ, ২,০৫০টি শব্দ আছে (খসড়া হিসাব দেখানো হলো)

এই দুই সহস্রাধিক শব্দকে আভিধানিক ধারানুযায়ী সাজিয়ে তাদের কোডিং প্রস্তুত এক বাইট (৮বিট) বাইনারী পদ্ধতিতে সম্ভব নয়। তাই ১৬বিট (ডবল বাইট) কোডিং পদ্ধতি ব্যবহার করা হয়েছে।এতে করে যে 2^{16} = (প্রায়) 64,000 address পাওয়া যায় তার একটা অংশ জুড়ে বাংলা এবং বাদবাকীতে ডাটাএন্ট্রি ফ্রেঞ্চ, জার্মান, ইটালিয়ান, স্প্যানিশ, রাশিয়ান এমনকি চাইনিজ ও জাপানিজ ভাষা প্রসেস করা সম্ভব হবে। বিভিন্ন গাণিতিক চিহ্ন, ছক তৈরীর সুবিধাও প্রদান করা যাবে। Sound coding-এর আরও একটি সুবিধা হলো Text to speech বাস্তবায়ন সম্ভব হবে।

উল্লেখ্য যে Double Byte Coding এর ফলে কমপিউটারের গতি (speed) এর তেমন উল্লেখযোগ্য হ্রাস হবে না। ASCII ও DBCS Code এর মধ্যে পার্থক্য করার ব্যবস্থা থাকবে। এবং শুধু সফ্টওয়্যারের পরিবর্তনের মাধ্যমেই ASCII সফ্টওয়্যারকে DBCS কোড গ্রহণযোগ্য করা সম্ভব তাই সবধরণের ইংরেজী সফ্টওয়্যারকে পরিবর্তন করে বাংলাকরণ করা যাবে।

DBCS কোডকে ডাটা কমিউনিকেশন (নেটওয়ার্কিং)এর কাজেও ব্যবহার করা যাবে, জাপান ও চীন যেমন করছে। এ ধরনের কোডিং-এর ফলে বাংলায় ডাটা কমিউনিকেশনের দক্ষতা উল্লেখযোগ্যভাবে বাড়বে (যদিও আপাতদৃষ্টিতে ১৬বিট কোডিং কমিউনিকেশনে ঝামেলা সৃষ্টি করবে বলে মনে হয়)। যেমন গ্রী-কে একটি ১৬বিট কোড প্রদান করেছি আমরা। এটাকে কমিউনিকেশনের জন্য ১৬টি Raw binary bit transfer করতে হবে। আর ৮ বিট কোডিং পদ্ধতিতে গ-এর জন্য ৮ বিট "্" এর জন্য ৮ বিট এবং ঈ-কার এর জন্য ৮ বিট অর্থাৎ মোট ২৪টি Raw bit transfer করতে হবে। কাজেই আমাদের কোডিং পদ্ধতিতে একই পরিমাণ বাংলা তথ্য ইন্টারচেঞ্জের জন্য যে সময় লাগবে তা অন্য যে কোন কোডিং পদ্ধতির চেয়ে কম (গড়পড়তা)। আমাদের এই সিস্টেমকে ইচ্ছে করলে স্বল্প খরচায় EPROM-এর মধ্যে ঢুকিয়ে Hardware Implementationও সম্ভব। এতে করে ফ্লপি ডিস্কের সংখ্যাও কমানো যাবে। তবে EPROM জাতীয় মেমরী ডিভাইস ব্যবহার না করেও পাঁচ/ছয়টি ফ্লপি ডিস্ক (360K)-এ এর কলেবর সীমিত রাখা যাবে। অবশ্যই এই কলেবরের ভিতরেই DBCS এর BIOS জাতীয় প্রোগ্রামসমূহও আছে।



চিত্র নং - ২

যে কোন মুহূর্তে একবার F10 চেপে ইংরেজী কীবোর্ডের ক্যারেক্টার সেট ব্যবহার করা যাবে এবং দ্বিতীয়বার F10 চেপে আবার বাংলাতে ফিরে আসা যাবে।

সর্টিং, ইনডেক্সিং ও ডাটাবেস ঃ

জুন ও জুলাই সংখ্যায় সিরাজুল হক সাহেব বাংলা সর্টিং ও ইনডেক্সিং এর-সমস্যা ও সম্ভাব্য সমাধান উল্লেখ করেছেন। আমাদের কোডিং পদ্ধতিতে সর্টিং ইনডেক্সিং আরও সুনিপুনভাবে করা সম্ভব। কারণ কা, কি, কো, কৌ ইত্যাদি এককভাবে সর্টিং, সার্চিং ও ইনডেক্সিংএর কী হিসাবে কাজ করবে, া, ি, ে, ৌ নিয়ে যে বিপত্তিগুলো সৃষ্টি হতো ওগুলো সমাধা হয়ে যাবে এমনকি রেফ (র্) ও কোন সমস্যা সৃষ্টি করবে না। কাজেই ডাটাবেস ম্যানেজমেন্ট আরও সুনিপুণ হবে অর্থাৎ কোন বাড়তি এলগরিদম বা প্রোগ্রামিং ছাড়াই (অর্থাৎ ইংরেজীর মতই) সার্চ, সর্ট, ইণ্ডেক্স সম্পন্ন হবে।

অপশনাল কেবলমাত্র ওয়ার্ড প্রসেসিং ফ্যাসিলিটি ঃ

(চিত্র নং ৩ এর কী বোর্ড, চিত্র ৪ এর উর্ধ্বাংশ দ্রঃ)

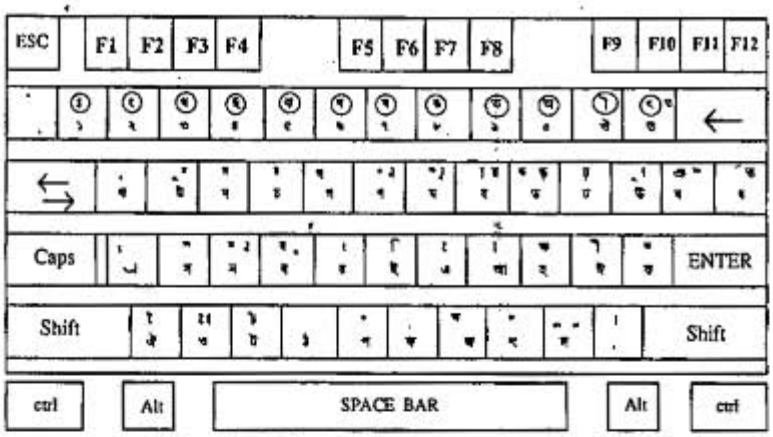
Alt-F* (Alt-F* বলতে Alt এর সাথে যে কোন একটি ফাংশন কী বুঝানো হয়েছে।) কী চেপে নিম্নরূপ একটি কী বোর্ডের মাধ্যমে সহজে ওয়ার্ড প্রসেসিং কাজ করার সুবিধা দেয়া যেতে পারে। আগেই বলা হয়েছে যে কোন ধরনের কী বোর্ড লে-আউট তৈরী সমস্যা নয়, সমস্যা হলো এভাবে যুক্তাক্ষরের চেহারা অবিকৃত রাখা, ব্যঞ্জনবর্ণের ঠিক নীচে ্ (র-ফলা) রাখা ইত্যাদি সমস্যার সমাধান করা যাবে না। সিরাজুল হক সাহেবের পদ্ধতিতে হয়ত সর্টিং, ইণ্ডেক্সিং করা যেতে পারে। এ ধরণের কোডিং ও কীবোর্ড লে আউটের একটি নমুনা নিম্নে দেয়া হোল(চিত্র নং -৩ ও চিত্র নং -৪ দেখুন)

উল্লেখ্য যে, বিভিন্ন ধরণের ইনপুট পদ্ধতির পরিবর্তন অথবা বিভিন্ন ভাষায় (যেমন জার্মান থেকে স্প্যানিস) ইনপুট পদ্ধতি পরিবর্তন আপনার সম্প্রতি ব্যবহৃত সফটওয়্যার (যেমন ওয়ার্ডস্টার বা ভিবেজ) স্ক্রীনকে বিনষ্ট করবে না অর্থাৎ আপনার কাজে কোন ব্যাঘাত না ঘটিয়ে আপনি ইচ্ছেমত সিস্টেমটি ব্যবহার করতে পারছেন। এতে Multilingual Text এডিট করা সুবিধাজনক হবে, যা ক্ষেত্রবিশেষে খুবই প্রয়োজনীয়।

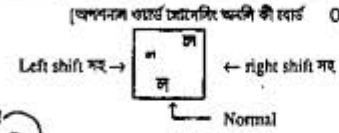
উপসংহার ঃ

যেহেতু আমরা, বাঙ্গালীরা এখানে সবাই ছাত্র, সময় এবং অর্থের সংকটে আছি তবুও নিষ্ঠার সাথে কাজ করলে আর মাস চারেকের ভিতরে মোটামুটি কাজ সমাপ্ত করতে পারব বলে আশা রাখি। তবে কোন সরকারী বা বেসরকারী প্রতিষ্ঠান বা কোন ব্যক্তি বিশেষের সুনজর হলে আমরা আর মাস খানেকের ভিতরেই সিস্টেমটির কাজ সমাপ্ত করতে পারব। BCTG এ ধরনের পৃষ্ঠপোষকতাকে স্বাগত জানাবে।

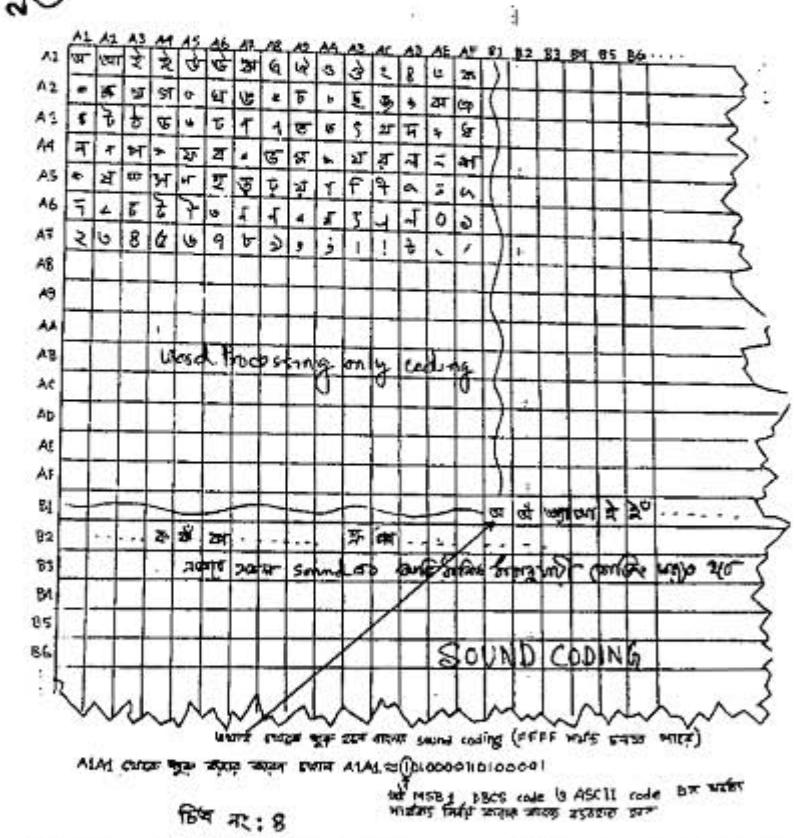
কমপিউটারে বাংলা প্রয়োগ সম্পর্কে যারা উৎসুক এবং কাজ করছেন তাদের সাথে মত বিনিময়ের আশা রাখি। এছাড়া যদি কোন সহৃদয় ভাষাবিদ বাংলা ভাষায় ব্যবহৃত Sound সমূহের (যেমন কা, অ, ম্ল, স্পি ইত্যাদি) frequency-র একটি তালিকা প্রস্তুতে



[অপশনাল ওয়ার্ড প্রসেসিং অনলি কী বোর্ড Optional Word Processsing only Key Board Lay-out]



[চিত্র নং : ৩]



চিত্র নং : ৪

সহযোগিতা করেন, তবে তাকে উপযুক্ত সম্মানী ও ধন্যবাদ দেয়া হবে। পদ্ধতিটি এরকম—সাহিত্যের একটি, বিজ্ঞানের একটি, বাণিজ্যের একটি বই নিন। বইগুলির প্রতিটি sound পড়ার সাথে সাথে sound-এর তালিকায় নির্দিষ্ট ঘরে এক যোগ করুন, এভাবে প্রতিটি sound-এর জন্য একটি সংখ্যা ওয়ালা টেবিল তৈরী হবে। সেটি পেলে আমাদের সিস্টেমের-এর উৎকর্ষ সাধন করা যাবে। অনেক উন্নত ভাষাতেই এধরণের Dictionary of Frequency জাতীয় বই আছে। একটা সমৃদ্ধ ভাষা হিসাবে বাংলা ভাষার গবেষণা কাজে এ ধরণের তালিকা সম্বলিত পুস্তকের গুরুত্ব অপরিসীম। উন্নত প্রযুক্তির স্বার্থে আমাদের মায়ের ভাষার গবেষণা আমাদেরই পবিত্র দায়িত্ব।

এছাড়া বাংলা ভাষার যুক্তাক্ষরের একটি তালিকা আমরা প্রস্তুত করেছি। আমাদের ধারণা দেশে এধরণের তালিকা বোধ হয় ভাষাবিদগণ করেছেন। তাই এধরণের একটি তালিকা পেলে আমরা আমাদের ভুলগুলি শুধরে নিতে পারব। এ ধরনের সাহায্যের জন্য আমরা সবিনয়ে কৃতজ্ঞতা স্বীকার করব।

যোগাযোগের ঠিকানা—

Md. Maruf Hasan
BCTG, Box I-41
Liu Xue Sheng Lou
Tsinghua University
Beijing 100084
P.R. China

উপসংহারে এটাই লিখতে চাই - কমপিউটার বাংলাকরণ ব্যাপারে Standardization যেন সরকার ও কর্তৃপক্ষ তড়িঘড়ি করে বা ঝোঁকের মাথায় চূড়ান্ত না করেন। এক্ষেত্রে প্রচুর গবেষনা ও টরাইঅল প্রয়োজন। বিভিন্ন দেশেই এ ধরনের কাজে প্রচুর সময় ও শ্রম ব্যয় করেছে এবং সবশেষে সবচেয়ে Prospective পদ্ধতিটি Standardize করে দেশের অর্থের সমূহ অপচয় রোধ করেছে। আমাদের দেশেও এটা হবার প্রত্যাশা সকল জনগণেরই। তাই ভাষাবিদ ও প্রযুক্তিবিদদের বেশী বেশী সেমিনার-আলোচনা-সমালোচনার সুযোগ দেয়া ও উদ্বুদ্ধ করা হোক। এতে করে নতুন নতুন পথের সন্ধান এমনকি একই ব্যক্তি বা গোষ্ঠীর থেকে সময়ের ব্যবধানে আলাদা পথের সন্ধান মিলবে। সবশেষে সবচেয়ে ভালোটি গ্রহণ ও অনুমোদন হোক এটাই কাম্য।

রেজিষ্ট্রি ডাকযোগে কমপিউটার জগৎ পেতে হলে
বার্ষিক ২০০ টাকা, ষান্মাসিক ১০৫ টাকা পাঠাতে হবে।
মানি অর্ডার, ড্রাফট পাঠাবার ঠিকানা ---
"মাসিক কমপিউটার জগৎ"
১৪৬/১, আজিমপুর রোড, ঢাকা-১২০৫
ফোন : ৫০৬৪ ৮৫

যারা লিখা পাঠাবেন
অবশ্যই কাগজের এক পার্শ্বে পরিষ্কার করে লিখবেন।
যথাসম্ভব বাংলা ভাষা ব্যবহার করবেন।
প্রকাশিত লেখার জন্য উপযুক্ত সম্মানী দেয়া হয়।