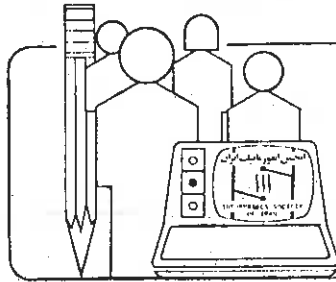


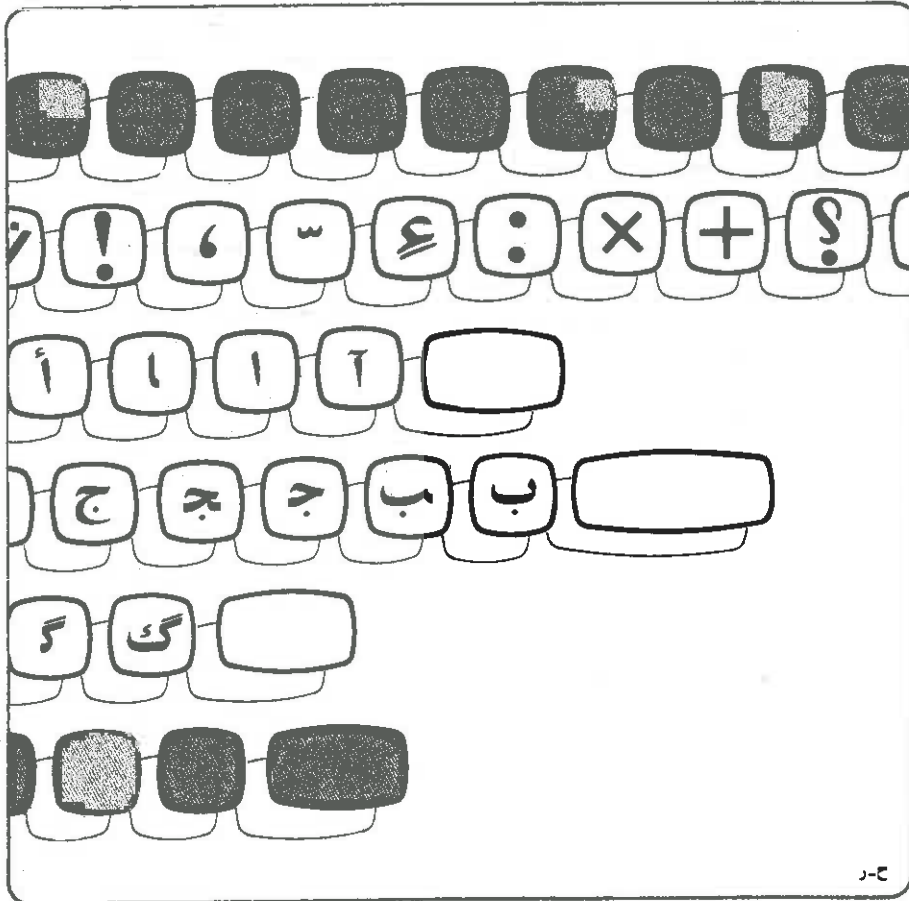
گزارش کامپیوتر

ماهنامه انجمن انفورماتیک ایران



سال هشتم - شماره ۲

شماره پیاپی ۷۶
اردیبهشت ۱۳۶۵



ج-ر



درباره این شماره

این شماره از نشریه گزارش کامپیوتر به مسئله کارکرد فارسی و دوزبانه در کامپیوترها اختصاص دارد. این مسئله در میان مسائل کامپیوتری کشور از اهمیت ویژه و اولسویت بالایی برخوردار بوده و هرکلیه جنبه‌های کاربرد کامپیوتر را بر آن به طور مستقیم و یا غیرمستقیم مؤثر است.

به علت کثرت مطالب، به ناچار بخش‌های متنوع نشریه که عمدتاً "به طور ثابت همه ماهه دنبال می‌شوند"، در این شماره حذف گردیده‌اند، تا آنجا که بخشی از مطالب مربوط به این بحث نیز در ماه آینده دنبال خواهند شد.

قبلاً در نشریه گزارش کامپیوتر مقالات متعددی در این زمینه درج گردیده است که در زیر به آنها اشاره می‌شود. علاقه‌مندان می‌توانند برای مطالعه بیشتر به این مقالات مراجعه نمایند. ذکر این نکته ضروری است که پارای از مطالب این شماره (به ویژه مطالب مربوط به مقالات ۱ تا ۵) قبلاً به نحوی از سبوی ارائه‌دهندگان نظریات مربوطه، درج گردیده است.

۱ - ترجمه متنهای ساده علمی از انگلیسی به فارسی به وسیله کامپیوتر

محمد میرزا عبدالحی، شماره ۸،

صفحات ۱۰ تا ۱۲، بهمن ۱۳۵۸

۲ - استانداردهای پیشنهادی برای کدبندی اطلاعات در فارسی

دکتر بهروز پیراهمی، شماره ۱۵،

صفحات ۴۳ تا ۴۹، شهریور ۱۳۵۹

۳ - طرح پیشنهادی زبان فارسی برای کامپیوتر

داریوش جوان، شماره ۲۲، صفحه ۱۲،

فروردین ۱۳۶۰

۴ - کامپیوتر و کشورهای عربی

دکتر پارتیک هال، ترجمه دکتر بهروز

پیراهمی، شماره ۲۵، صفحات ۱۶ و ۱۵،

تیر ۱۳۶۰

بقیه در صفحه ۲

در این شماره

۲ اخبار انجمن.....

۹ نقطه شروع مبتدیان - کارکرد فارسی و دوزبانه در کامپیوتر (حمیدرضا رهبر).....

مقاله ۱..... پیشنهادی کدبندی اطلاعات در فارسی (کمیته مطالعاتی

استاندارد اطلاعات در فارسی - سازمان برنامه و بودجه).....

مقاله ۲..... مختصری درباره روش استاندارد برای توسعه کد ISO (کمیته مطالعاتی

استاندارد اطلاعات در فارسی - سازمان برنامه و بودجه).....

مقاله ۳..... مختصری درباره کدگذاری برای الفبای فارسی (کمیته مطالعاتی

استاندارد اطلاعات در فارسی - سازمان برنامه و بودجه).....

مقاله ۴..... نکاتی درباره نمایش علائم فارسی در دستگاههای خروجی (کمیته مطالعاتی

استاندارد اطلاعات در فارسی - سازمان برنامه و بودجه).....

مقاله ۵..... نکاتی درباره طرح صفحه کلیدهای فارسی و دوزبانه (کمیته مطالعاتی

استاندارد اطلاعات در فارسی - سازمان برنامه و بودجه).....

مقاله ۶..... کدبندی برای تبادل اطلاعات در فارسی (دکتر محمد صنعتی).....

مقاله ۷..... پردازش کدگذاری: تعیین کننده مهم در تدوین استاندارد در تبادل اطلاعات

در فارسی (قاسم جابری پور).....

۱۹.....

بقیه سرمقاله

۵ - شناسائی متون چاپی فارسی توسط کامپیوتر
دکتر بهروز پیرهای ، شماره ۲۷ ،
صفحات ۷ تا ۱۱ ، شهریور ۱۳۶۰

۶ - مسائل مربوط به زبان و خط در کاربردهای
کامپیوتر

دکتر بهروز پیرهای ، شماره ۲۹ ،
صفحات ۵ تا ۱۲ ، آبان ۱۳۶۰

۷ - چاپ خط درشت فارسی توسط کامپیوتر
پرویز دوستی ، شماره ۳۲ ،
صفحات ۱۸ و ۱۹ ، بهمن و اسفند ۱۳۶۰

۸ - پایانه فارسی با حداقل تعداد کلید
منصور جمزاده ، شماره ۳۲ ،
صفحات ۶ و ۷ ، اردیبهشت ۱۳۶۱

۹ - طرح چندزبانانه کردن چاپگرهای کامپیوتری
" ال ای ۳۶ " و " ال ای ۱۸۰ "

حسین فیروزبخت و امیر اسعد انزانی
شماره ۳۵ ، صفحات ۱۴ و ۱۵ ، خرداد ۱۳۶۱

۱۰ - تغییر خط فارسی ماشینهای باروز سری
ال ۹۰۰۰

محمد رضا حبیبیان ، شماره ۴۰ ،
صفحات ۵ و ۶ ، آذرود ۱۳۶۱

۱۱ - درباره علامتی برای نمایش واحد پول
ایران

دکتر بهروز پیرهای ، شماره ۵۲ ،
صفحه ۲۲ ، فروردین و اردیبهشت ۱۳۶۳

۱۲ - فارسی کردن زبانهای برنامه سازی سطح
بالا

قاسم جابری پور ، شماره ۶۸ ،
صفحات ۵ تا ۷ ، شهریور ۱۳۶۳

۱۳ - یک راه حل ساده برای مرتب سازی
اطلاعات فارسی

قاسم جابری پور ، شماره ۷۴ ،
صفحات ۳ تا ۸ ، اسفند ۱۳۶۴

سردبیر



اعضاء

مساء سفانه تا آخرین لحظاتی که نشریه
برای چاپ آماده می شد ، اما را غای انجمن
به دست کمیته انتشارات نرسید . در شماره آینده
تغییرات اعضا در هر دو ماه ، درج خواهد شد .

در این جا با ردیگر توجه علاقه مندان عضویت
و نیز اعضای انجمن را به افزایش حق عضویتها ،
جلب می کنیم تا در هنگام عضویت ویا تمدید
عضویت آنها را رعایت نمایند .

ضمناً " همان گونه که در اخبار انجمن نیز
قید شده است ، فرم تقاضای عضویت که موجودی
آن به اتمام رسیده بود تجدید چاپ شده است
که علاقه مندان می توانند از طریق کمیته روابط
عمومی انجمن آن را تهیه نمایند .

حق عضویت سالانه اعضا : [عبادی ۲۵۰۰ ریال
(شامل اشتراک ماهنامه) وابسته ۲۵۰۰ "]
[دانشجوی ۱۰۰۰ "]

حق اشتراک یک ساله گزارش کامپیوتر ۱۰۰۰ "]
هزینه پست برای اعضا با مشترکین
مقیم خارج از کشور ۱۰۰۰ "]

لطفاً " رسید پرداخت وجه به حساب جاری
شماره ۱۱۱۷ انجمن نزد بانک صادرات شعبه
دانشگاه صنعتی شریف ، تهران (شعبه شماره
۱۶۵۲) را به درخواست خود ضمیمه کنید .
پرداخت وجه از طریق هر یک از شعب بانک
صادرات در تهران و شهرستانها ممکن است .
برای دریافت بروشور حاوی فعالیتها
انجمن و تقاضای عضویت ، با ما
مکاتبه کنید .

اخبار انجمن



برگزاری موفقیت آمیز نخستین سمپوزیوم

نوروسیکولوژی ایران

نخستین سمپوزیوم نوروسیکولوژی از
۶۵/۲/۱۷ الی ۶۵/۲/۱۷ در محل دانشکده علوم
توانبخشی برگزار شد که طی آن سخنرانان متعدد
مطالب گوناگونی را در زمینه نوروسیکولوژی
ارائه نمودند . دوشن از اعضا انجمن
انفورماتیک ایران از جمله سخنرانان این
سمپوزیوم بودند که سخنرانیهای آنان با
استقبال مواجه گردید .
گزارش مبسوطتر برگزاری این سمپوزیوم
در شماره آینده درج خواهد گردید .

تجدید چاپ فرم عضویت

فرم تقاضای عضویت در انجمن با اندکی
تغییرات نسبت به فرم سابق تجدید چاپ
گردید . بدینوسیله از علاقه مندان می خواهیم
جدیداً " به عضویت انجمن در آیند درخواست
می شود فرم مزبور را از طریق انجمن تهیه
نمایند .

افزایش هزینه های چاپ ماهنامه

نظریه این که انجمن کاغذ مورد نیاز
خود را از منابع آزاد تهیه می نماید ، رشد
هزینه های کاغذ ، مخارج چاپ ماهنامه را هر
ماه افزایش می بخشد . هزینه کاغذ در این
ماه نسبت به ماه قبل حدود دو برابر نسبت
به سال قبل همین موقع حدود چهار برابر
گردیده است . البته در جهت تأمین کاغذ به
صورت دولتی از سوی کمیته انتشارات اقداماتی
در دست انجام است ولی هنوز نتیجه قطعی
حاصل نشده است .

برگزاری دوره های آموزشی

از سوی کمیته آموزش انجمن دوره های
آموزشی ای برنا مدیری مقدماتی شده است .
ولیکن متأسفانه به دلیل عدم وجود امکانات
از قبیل محل برگزاری ، تاکنون هیچ یک به
مرحله نهائی نرسیده است . تأمین امکانات
برگزاری دوره ها از سوی مسئولین انجمن در
دست پیگیری است که در صورت مثبت بودن نتایج
اعضای انجمن در جریان امور قرار خواهند گرفت .

برای تماس

برای تماس با انجمن از نشانی یا شماره تلفنهای زیر استفاده کنید :
نشانی : انجمن انفورماتیک ایران ، صندوق پستی ۱۹۹۶-۱۴۱۵۵ تهران

شماره تلفنهای مسئولان انجمن

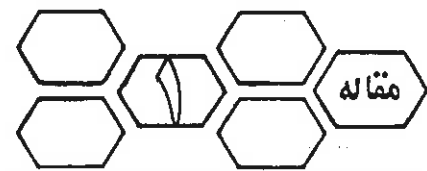
رئیس آقای ابراهیم تقی زاده ، متابع داخلی ۲۲۷ - ۴ - ۲۲۹۰۷۱
نائب رئیس آقای محمد میرزا عبداللهی ۴۲۷۲۲۲۷
دبیر خانم لیلی حاجی اسماعیلی ۸۹۱۲۲۹ و ۸۹۱۴۲۹
خزانه دار آقای دکتر محسن رستمی ۸۹۱۲۹۳ و ۸۹۱۴۲۹

شماره تلفنهای سرپرستان کمیته ها

انتشارات آقای حمیدرضا رهبر ۸۹۱۲۹۳ و ۸۹۱۴۲۹
آموزش آقای دکتر محسن رستمی ۸۹۱۲۹۳ و ۸۹۱۴۲۹
فعالیت های علمی و فنی آقای محمد میرزا عبداللهی ۸۹۳۲۵۱
عضویت آقای مهدی فلاحتی ۸۹۱۲۹۳ و ۸۹۱۴۲۹
روابط عمومی آقای ابراهیم تقی زاده ، متابع داخلی ۲۲۷ - ۴ - ۲۲۹۰۷۱

دیگر اعضای هیأت اجرایی

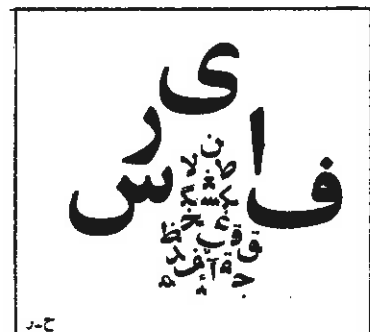
عضو اصلی آقای عبداللہ کسرائی ۸۹۱۴۲۹ و ۸۹۱۲۹۳
عضو علی البدل آقای مهرداد دمیرنیا ۸۹۱۴۲۹ و ۸۹۱۲۹۳
عضو علی البدل آقای داریوش جوان تبریزی ۲۹۲۳۰۴



پیشنهادهای کد استاندارد تبادل اطلاعات در فارسی

کمیته مطالعاتی استاندارد
تبادل اطلاعات در فارسی

سازمان برنامه و بودجه



کارکرد دوزبانه / کارکرد فارسی / جدول
رمزبندیها / استاندارد سازی / پسردانش
اطلاعات / روش نگارشی

پادداشت سردبیر

این مقاله و نیز چهار مقاله بعدی در این شماره، حاصل کدهای تالشی "کمیته مطالعاتی استاندارد تبادل اطلاعات در فارسی" است. این کمیته در سال ۱۳۵۶ بنا به دعوت معاونت انفورماتیک سازمان برنامه و بودجه، متشکل از نمایندگان سازمانها و مراکز مختلف، سازمانهای ذینفع و دانشگاهها به منظور مطالعه و پیشنهاد نمودن کدهای استاندارد تبادل اطلاعات در فارسی تشکیل گردید و از تاریخ ۱۳۵۶/۱۱/۲۹ کار خود را آغاز نمود.

این مقالات قبلاً در یک مجموعه به نام "پیشنهادهای کدهای تبادل اطلاعات در فارسی" از سوی دفتر هماهنگی و ترویج مسأله‌ها، انتشارات انفورماتیک سازمان برنامه و بودجه به چاپ رسیده است که اکنون به درخواست شورای عالی انفورماتیک و به دلیل عدم دسترسی بسیاری از افراد جامعه انفورماتیک ایران به آن و نیز به منظور آشنایی افکار و دیگرخواهانندگان گزارش کامپیوتر با فعالیتهای انجام شده در زمینه کدگذاری زبان فارسی، بر روی کامپیوترها، در اینجا با اندکی جرح و تعدیل به چاپ می‌رسد. در ادامه، این مجموعه از اهداف، اهداف و کمیته‌های آن، نام اعضای کمیته، و نیز نامهای افراد و کادر درگیر در تهیه گزارش همکاری مؤثر آنها، تشکر شده است.

چون هر نوع استاندارد تهیه شده باید برای مدت طولانی مورد استفاده قرار گیرد و از طرف دیگر تکنولوژی کامپیوتر نیز به سرعت در حال تغییر و تحول است، اهمیت عدم وابستگی استاندارد بین‌المللی به تکنولوژی‌های خاص، آشکار می‌شود.

مقدمه

اهمیت استانداردهای برای ایجاد همسازی (Compatibility) بین سیستمهای مختلف محتاج به تائید نیست. در مورد کاربردهای کامپیوتری و پردازش اطلاعات، وجود استانداردها از نقطه نظرهای زیر مورد لزوم است:

۱- از نظر سازندگان کامپیوتر: برای حذف مطالعات تکراری هنگام طرح سیستمهای جدید، و تسهیل شدن بازاریابی و فروش و ایجاد ثبات و اطمینان برای سرمایه‌گذاری در ارائه محصولات متنوع.

۲- از نظر استفاده‌کنندگان: برای ایجاد همسازی بین سیستمهای مختلف، آزادی بیشتر در انتخاب اجزای مورد نیاز و هر سیستم و امکان اشتراک مساعی از طریق استفاده از نتایج کار دیگران.

۳- از نظر متخصصین کامپیوتر: برای ایجاد یکپارچگی در سیستمها و روشهای استفاده از آنها، امکان استفاده از محصولات خارجی و عدم نیاز به مطالعه خاص منحصر بفرد در هر سیستم جدید.

بطور کلی در هر زمینه‌ای پیدایش استانداردها به دنبال بلوغ نسبی در کاربردهای مختلف و ایجاد آگاهی کافی به مسائل موجود و به منظور جلوگیری از هرج و مرج و تضاد منابع و امکانات شروع می‌شود. در ایران نیز با توجه به توسعه کارهای انفورماتیک در سالهای اخیر (۱) لزوم تهیه چنین استانداردهایی به صورتهای مختلف تأکید شده است (مثلاً (۲)). منافع حاصل از تبعیت از چنین استانداردهایی به دو دسته قابل تقسیم هستند:

۱- منافع ملی: نشان دادن رشد تکنولوژی انفورماتیک و کاربردهای آن در کشور، تقلیل درجه وابستگی به سازندگان سیستمهای کامپیوتری بخصوصی با به‌گذاری منابع داخلی انفورماتیک به طریق صحیح.

۲- منافع اقتصادی: عدم نیاز به مطالعات تکراری و برنامه‌نویسی مجدد در سطح سیستمهای جدید، جلب سازندگان بیشتر به بازار ایران و ایجاد رقابت بین آنها، امکان انتخاب با صرفه‌ترین اجزاء در

سازمانها و نهادهای دولتی و خصوصی. اما حائزترین استاندارد برای شروع کار در این زمینه، استاندارد بین‌المللی اطلاعات (ISO) است.

با توجه به مشترک بودن الفبای ما با زبان عربی چنین کد استاندارد، در صورت توافق کشورهای عربی زبان، در ناحیه وسیعی از دنیا قابل استفاده خواهد بود. استفاده از یک کد استاندارد در کلیه مواردی که اطلاعات فارسی بین سیستمهای مختلف رد و بدل می‌شود، ما این امکان را خواهد داد که بدون نیاز به دانستن مشخصات سیستمهای فرستنده و گیرنده اطلاعات بتوانیم اطلاعات ارسالی را به صورت واحدی تعبیر و تفسیر کنیم.

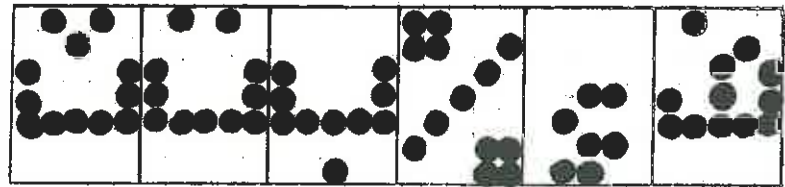
ایجاد این نوع ارتباط بین سیستمهای مختلف در حال حاضر توسط برنامه‌های مترجم مخصوص صورت می‌گیرد. بدیهی است که با ازدیاد روزافزون انواع سیستمهای مختلف، این نحوه عمل دارای بازده بسیار کمی است و باعث اتلاف نیروی انسانی می‌شود زیرا هر سیستم باید برنامه مترجم مخصوص به آن تهیه شود. واضح است که در کلیه مواردی که اطلاعات صرفاً در داخل یک سیستم (که نیازی نیز به تبادل اطلاعات به سایر سیستمهای دیگر ندارد) رد و بدل می‌شوند تبعیت از یک کد استاندارد لازم نیست، اگرچه استفاده از چنین کدی توسعه و ارتباطهای بعدی سیستم را با سیستمهای دیگر ساده‌تر می‌کند.

این نوع کدهای استاندارد برای تبادل اطلاعات تا سالهاست که در کشورهای غربی به وجود آمده‌اند. مهمترین این کدها از نظر وسعت کاربرد و شناسایی در سطح بین‌المللی کد استاندارد ISO - IRV (شکل ۱) است (۳) که فقط با دو اختلاف (در علامت و جدولی ۲/۴ و خط بالای حروف ۷/۱۴) همان کد استاندارد ASCII - می‌باشد (۴). به همین جهت پیشنهاد داده شده در این گزارش براساس توسعه همین کد و با در نظر گرفتن خط مشی تعیین شده برای توسعه آن به صورت استاندارد (۵) تهیه شده است. مقاله ۲ حاوی شرح مختصری از روش استاندارد برای توسعه کدهای ISO می‌باشد.

با توجه به داشتن کدهای استاندارد نمودن استانداردها در تمام زمینه‌ها با اشکالاتی روبرو می‌شود که از آن جمله می‌توان وجود سیستمهای غیر استاندارد و سرمایه‌گذاریهای وسیع در ایجاد این سیستمها و تکنولوژیهای وابسته به آنها را نام برد. به این جهت است که اغلب کمیته‌های پیشنهاد کننده استانداردها را در مجبور می‌شوند مابین ثابت شده علمی و عملی را به نفع کاهش زبانهای اقتصادی، نادیده گرفته و استاندارد را وضع نمایند که بدون اشکال زیاد مورد قبول اکثریت

میگیرند همگی بر اساس زبانهای غربی و خط لاتین طرح ریزی شده اند. به عنوان مثال دستگاه چاپ زنجیری (Chain Printer) که برای خط لاتین بسیار مناسب است، به علت وجود فاصله بین چکشها نمی تواند در مورد خط فارسی حروف مجاور را به یکدیگر متصل نماید. به عنوان مثال دیگر می توان نامناسب بودن ماتریسهای نقطه ای متداول 7×5 با 9×7 را برای نمایش حروف فارسی در نظر گرفت. با توجه به این نوع محدودیتهای تکنولوژی موجود، سازندگان مختلف راههای موقتی به خصوصی برای بعضی از اشکالات فوق الذکر یافته اند که تنوع این راه حلها خود را در برابری آنها به تکنولوژیهای خاص است. چون هر نوع استاندارد تهیه شده با بد برای مدت طولانی مورد استفاده قرار گیرد و از طرف دیگر تکنولوژی کامپیوتر نیز به سرعت در حال تغییر و تحول است، اهمیت عدم وابستگی استاندارد پیشنهادی به تکنولوژی (های) خاص آشکار می شود.

"کمیته مطالعاتی استاندارد ملی اطلاعات در فارسی پس از بحثهای اولیه مستقد گردید که



کدام استاندارد ISO-IRV نظر و سمیت کاربردها و عملیاتی در سطح بین المللی مهمترین استاندارد در این زمینه است که فقط با دو اختلاف همان استاندارد ASCII است.

- ۱ - مرتب کردن اطلاعات با توجه به اشکال مختلف هر حرف در متنها.
- ۲ - جدا کردن و شناسایی کلمات به منظور پردازش متون فارسی.
- دستگاهها و برنامه های که در حال حاضر برای کار با اطلاعات فارسی مورد استفاده قرار

جا معما ستفا ده کنندگان از آن قرار گیرد. به عنوان مثال، مطالعه نشان داده است که استفاده از صفحه کلیدی که کریستوفر شولز (C. Sholes) در سال ۱۸۷۲ میلادی برای ماشینهای تحریر پیشنهاد نمود در رابطه با سیستمهای کامپیوتری بیشتر به خاطر آشنایی میلیونها نفر با ترتیب قرار گرفتن علائم روی آن بوده است (۶). عدم وجود این عامل ممکن بود باعث پیدایش استاندارد نامناسبتری برای این منظور میشد. به همین جهت، "کمیته مطالعاتی استاندارد ملی اطلاعات در فارسی" لزوم سرعت عمل را در قبول استاندارد رده های در این زمینه تا کید میکند. سرما به گذار یهای وسیع فعلی به بروی سیستمهای غیر استاندارد باعث خواهد شد که وضع نمودن استاندارد در آینده به مراتب مشکلتر و محتاج به منظور نمودن عوامل بیشتری در رابطه با سیستمهای موجود در آن زمان باشد.

مسائل خاص زبان و خط فارسی

بر اثر تجربیات حاصل از کار برده های کامپیوتری در ایران و برخی مطالعات انجام شده (مثلا (۷) و (۸) اشکالات ناشی از زبان و خط فارسی در رابطه با این کار برده ها تا حدود زیادی شناخته شده اند. مهمترین اشکالات عبارتند از:

- ۱ - تعداد نسبتاً زیاد حروف در الفبای فارسی با در نظر گرفتن گونه های مختلف نوشتن هر حرف.
- ۲ - اهمیت زیبایی در خط فارسی و تا کید اجتماع بر روی آن به علت سوابق تاریخی.
- ۳ - نوشتن اطلاعات عددی و غیر عددی در جهت های مختلف.
- ۴ - لزوم اتمال بسیاری از حروف به یکدیگر در متون فارسی.
- ۵ - اختلاف اندازه های هندسی و به خصوص عرض در خطهای زیبا خوانا.
- ۶ - وابستگی شکل ظاهری حروف در متن به حروف مجاور.
- ۷ - اختلاف جزئی در شکل ظاهری بعضی از حروف مانند "ب" و "پ".
- ۸ - لزوم در نظر گرفتن اطلاعات مخلوط فارسی و لاتین در بسیاری از کار برده ها.

شماره ستون

(b₇ b₆ b₅) ←

شماره سطر

(b₄ b₃ b₂ b₁) ↓

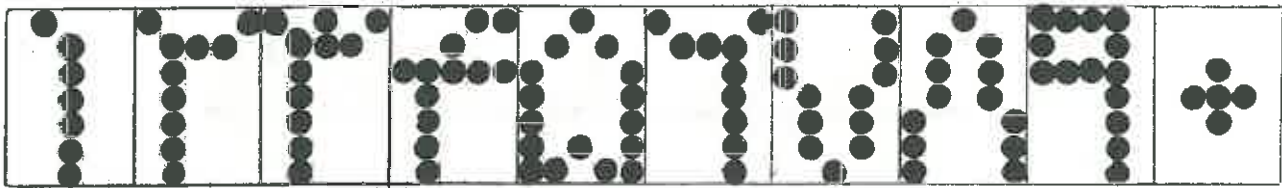
	0	1	2	3	4	5	6	7
0	NUL	DLE	SP	0	@	P	`	p
1	SOH	DC1	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3	#	3	C	S	c	s
4	EOT	DC4	□	4	D	T	d	t
5	-ENQ	NAK	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(	8	H	X	h	x
9	HT	EM	)	9	I	Y	i	y
10	LF	SUB	*	:	J	Z	j	z
11	VT	ESC	+	;	K	[	k	{
12	FF	FS	,	<	L	\	l	
13	CR	GS	~	=	M	]	m	}
14	SO	RS	.	>	N	^	n	—
15	SI	US	/	?	O	_	o	DEL

علائم کنترل

۹۴ نمایشی با ضاعه علائم sp و DEL

شکل ۱: کدام استاندارد ISO-IRV برای تبادل اطلاعات

نظر به کاربرد نسبتاً زیاد علامت "\$" در ایران، توصیه میشود که این علامت در ستون ۲ سطر ۴ بعنوان علامت پولی مورد استفاده قرار گیرد.



به اقتضای امکانات و محدودیت‌های سخت افزار مربوطه از کدهای اطلاعاتی مخصوص به خود استفاده خواهند نمود، چون در اکثر این کدها رابطه یک بیک (از نظر تعداد مکان‌های اشغال شده) بین علائم موجود در کد و علائم چاپی یا نمایی موجود است، تبدیل اطلاعات ارائه شده در این کدها به تک نمایی به علت لزوم استفاده از علائم‌های PS و PC با عتادها دطول بعضی از قطعات اطلاعاتی (فیلدها) خواهد شد و با این جهت در صورتی که حفظ اطلاعات به صورت کامل مورد نظر باشد، لزوماً قالب (Format) فایلها باید تغییر نماید. نظیر همین مسئله در دستگا‌های چاپ و نمایش پیش می‌آید که وجود علائم PS و PC (که مکانی را در چاپ و نمایش اشغال نمی‌کنند) ممکن است باعث بهم خوردن فرم جدول بندی در صفحه خروجی شوند. اگرچه ممکن است بتوان این نوع اشکالات را با اعمال تغییرات مناسبی در نرم افزار سیستمها برطرف نمود، وابستگی نسبتاً زیاد ما به نرم افزار ارائه شده توسط سازندگان، تغییرات و اصلاحات مداوم در نرم افزار سیستمها و نیروی انسانی لازم برای اعمال این تغییرات این کار را غیر عملی و با حداقل غیر اقتصادی می‌کند.

به همین جهت "گمته مطالعاتی استاندارد تبادل اطلاعات در فارسی" پس از بحث‌های مفصل به این نتیجه رسید که با توجه به شرایط و محدودیت‌های فعلی، حفظ ارتباط یک بیک بین علائم موجود در کد استاندارد و علائم چاپی یا نمایی در دستگا‌های خروجی با پند نسبت به سایر عوامل از اولویت بیشتری برخوردار گردد.

کد استاندارد پیشنهادی برای علائم

فارسی

نظر به متداول بودن کد استاندارد ISO-IRV، استاندارد پیشنهادی براساس توسعه آن از طریق روش پیشنهادی ISO (۵) صورت گرفته است. مخصری از این روش در مقاله ۲ ارائه شده است. شکل ۲ مجموعه G1 فارسی از کد توسعه یافته را نشان می‌دهد. مکان‌های که در انتخاب حروف و علائم محل قرار گرفتن آنها در جدول رعایت شده اند عبارتند از:

- ۱ - ارتباط یک به یک بین علائم موجود در کد و علائم چاپی یا نمایی.
- ۲ - دربرگرفتن کلیه علائم متداول خاص متون فارسی.
- ۳ - حفظ ترتیب الفبایی علائم به منظور ایجاد سهولت در مرتب کردن اطلاعات.

برای شروع کار اساسیترین استاندارد مربوط به کد نمایی علائم فارسی است.

شماره ستون

	2	3	4	5	6	7
0	SP	۵	×	۷	—	ل
1	!	۱	ۛ	خ	ض	ا
2	"	۲	ۛ	خ	ط	م
3	#	۳	ۛ	د	ط	م
4	R	۴	ۛ	ذ	ظ	ن
5	%	۵	ۛ	ر	ظ	ز
6	÷	۶	ۛ	ز	ع	و
7	'	۷	ۛ	ز	ع	ه
8	(	۸	ۛ	س	غ	ه
9	)	۹	ۛ	ش	غ	ی
10	*	:	ۛ	ص	ف	ی
11	+	؛	ج	ـ	ق	ـ
12	,	«	ج	ـ	ق	ا
13	-	=	ج	ـ	ک	۶
14	.	»	ج	ـ	ک	۷
15	/	؟	ح	ـ	لا	DEL

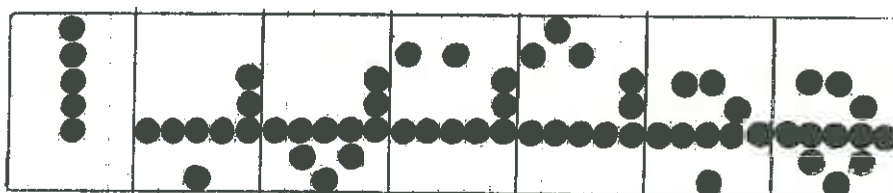
شکل ۲:

مجموعه علائم G1 استاندارد برای الفبای فارسی

یکی از راه‌های تأمین استقلال استاندارد پیشنهادی از تکنولوژی‌های خاص استفاده از کد تک نمایی برای نمایش حروف الفبای فارسی است. در این کد که برای اولین بار توسط "حیدر" (۹) پیشنهاد گردید و اصول آن در مقاله ۳ تشریح شده است، هریک از حروف الفبای فارسی فقط به یک صورت نمایی داده می‌شود و فرم‌های آن در متون فارسی توسط دستگا‌ها پراکنده با توجه به حروف مجاور تعیین می‌شود. مزایای استفاده از کد تک نمایی به اختصار عبارتند از:

- ۱ - کم شدن تعداد علائم (اندازه کد) و آسانی بیشتر در استفاده از علائم مخصوص در صورت در نظر گرفتن کدی با اندازه ثابت.
- ۲ - استقلال کامل روش نمایی داخلی اطلاعات از روش‌های مورد استفاده در دستگا‌های خروجی به منظور تولید خط فارسی زیبا و خوانا.
- ۳ - از بین بردن وابستگی اطلاعات به اثر کم شدن تعداد علائم و در نتیجه استفاده کمتر از کلید مبدل (Shift).
- ۴ - ایجاد سهولت در پردازش اطلاعات به خاطر این که نمایی داخلی هریک از حروف الفباء مستقل از متن است.
- در مقابل مزایای فوق می‌توان معایب زیر را برای کد تک نمایی ذکر نمود:
- ۱ - اتلاف فضای حافظه و پهنای آمدن یا زده کار در انتقال اطلاعات به خاطر لزوم استفاده از علائم‌های فاصله مجازی (Pseudo-Space) و اتصال مجازی (Pseudo-Connection) (۱۰).
- ۲ - لزوم وجود امکان پردازش اطلاعات در دستگا‌های ورودی و خروجی و وابستگی‌های کنترل کننده آنها.
- ۳ - عدم آشنایی پرسنل موجود با مفاهیم مربوط به کد تک نمایی.
- ۴ - عدم تناظر یک بیک بین علائم موجود در کد و علائم چاپی یا نمایی در دستگا‌های خروجی.

اگرچه با توجه به گرایش‌های نوین در طراحی سیستم‌های کامپیوتری و کاهش روزافزون هزینه سخت افزار، اشکالات اول و دوم اهمیت نسبتاً کمی دارند و در آینده نزدیک به کلی منتهی خواهد شد. اشکال سوم نیز نظیر بک‌موند تعداد پرسنل موجود و مزایای درازمدت استفاده از کد تک نمایی قابل اغماض است (۱۰). ولی اشکال چهارم در عمل ممانعتی بوجود می‌آورد. علت اصلی آن است که کار بر روی کد استاندارد (حداقل در آینده نزدیک) فقط محدود به تبادل اطلاعات خواهد بود و سازندگان مختلف سیستم‌های کامپیوتری هریک



باید در زمینه وضع و قبول استاندارد نهایی سرعت عمل به خرج داد زیرا سرمایه گذاریهای وسیع فعلی بر روی سیستمهای غیر استاندارد باعث خواهد شد که در آینده این امر به مراتب مشکلتر گردد.

۴ - سهولت تشخیص ارقام وحروف وفشرده کردن ویا زکرن اطلاعات عددی .

۵ - دربرگرفتن کلیه علائم مخصوصی که در کاربردهای متداول لازمند .

۶ - حفظ تناظر محل قرار گرفتن علائم معادل یا مشابه در جدول فارسی ولاتین تا حد امکان، اهمیت ارتباط یک به یک بین علائم موجود در کد و علائم چاپی یا نمایشی را قبلا تشریح کردیم. در زیر بقیه این نکات را با تفصیل بیشتر مورد بررسی قرار میدهم.

در کد پیشنهادی سعی شده است که کلیه علائم لازم برای نمایش اطلاعات فارسی گنجانده شوند. به این جهت تمام علائم نقطه گذاری فارسی شامل گیومه های باز و بسته، چهار عمل اصلی و سایر علائم ریاضی، بعضی از علائم مخصوص لاتین که در فارسی نیز کاربرد دارند (مثلا عنصر ستون ۲ سطر ۷ جدول که برای جدا کردن اعداد و یا به عنوان "الف مقصوره" قابل استفاده است) علائم تجاری ریا ل و درم (ستون ۲ سطرهای ۴ و ۵)، حرکات و سایر علائم مخصوص خط فارسی و یا لایحه خط تیره برای کشیدن حروف در جدول کد وجود دارند. در مورد حرکات و سایر علائم مخصوص خط فارسی، روش ساده ای برای استفاده در دستگا های چاپ یا نمایش معمولی (بدون قابلیت چاپ مضاعف یا روی هم) وجود دارد که در مثالهای زیر ارائه شده است (۱):

م ت ک ل م ن ا
ا ر د ک م ا و ا

تصمیم گیری در مورد فرم چاپی علائم مخصوص فوق با توجه به حرف قبلی امکان پذیر است. این موضوع که در مورد حروف الفباء (مثلا حروف "ع") نیز صدق میکند اگرچه سبب میشود که برای بدست آوردن خط زیبا و کاملاً خوانا، دستگا های خروجی با قابلیت پردازش اطلاعات مورد نیاز با شنودلی حداقل در مقایسه با کد تک نما دی این مزیت را دارنده تصمیم گیری در مورد شکل حروف و علائم به حروف یا علائم بعدی بستگی ندارد و به این جهت هیچ نوع تاخیری در ظهور علائم وارد شده بر روی دستگا و چاپ یا نمایش لازم نیست. به

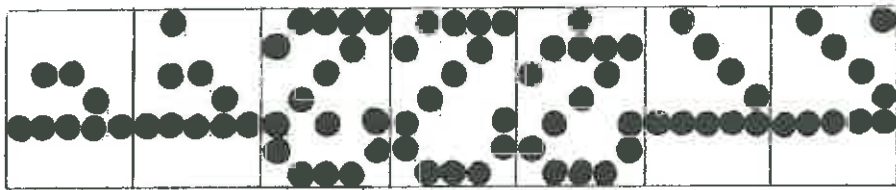
علاوه بر صورت قبول کردن خطی نه چندان زیبا، میتوان از دستگا های خروجی بسیار ساده و ارزان قیمت نیز برای تولید خط فارسی استفاده نمود. مسئله مرتب کردن اطلاعات در فارسی برای کلیه کسانی که با کاربردهای کامپیوتری در ایران سروکار داشته اند مطرح بوده است. متأسفانه به خاطر خواص خط فارسی، حل این مسئله صرفاً توسط کدگذاری مناسب برای علائم ممکن نیست (۸) به این جهت در مورد ترتیب الفبائی علائم فقط سعی شده است که ارتباط طبیعی آنها حفظ شود و علائم "دنباله تخت" و "دنباله گود" (ستون ۴ سطرهای ۱ و ۲) در محلها یی قرار گیرند که حداً قدر اشکال را تولید نمایند. در رابطه با همین موضوع، فرم ترکیبی "لا" به خاطر نقشی که در تولید خط فارسی زیبا دارد و نظر به اهمیت ارتباط یک به یک بین علائم موجود در کد و علائم چاپی یا نمایشی (که قبلاً به آن اشاره شد)، به صورت یک علامت جداگانه در جدول کد در نظر گرفته شده است. توصیه میشود که به خاطر اشکالاتی که استفاده از "لا" در مرتب کردن اطلاعات تولید میکند، حتی الامکان به جای آن از دو حرف "ل" و "ا" استفاده شود.

برای سهولت در تشخیص ارقام وفشرده کردن و با زکرن اطلاعات عددی، ارقام صفرتا ۹ مطابق با استاندارد ISO در محلها یی قرار گرفته اند که چهار بیت سمت راست کدهای رقم، مقدار عددی آنرا در مبنای ۲ مشخص میکند. در مورد علائم الفبائی تمام اعداد علائم لازم مساوی ۵۲ میباشد که مکانهای نظیر حروف بزرگ و کوچک لاتین را در جدول کد استاندارد ISO اشغال کرده اند. به این ترتیب کلیه آزمایشهای کد بر بنا مهای سیستم ویا بر بنا مهای مترجم زبانهای سطح بالا در مورد عددی یا حرفی بودن اطلاعات به کار میروند و عیناً وبدون هیچگونه تغییری در مورد کد فارسی پیشنهادی نیز قابل اجرا خواهند بود.

در مورد علائم مخصوص سعی شده است که کلیه علائم لازم برای کاربردهای متداول در جدول کد گنجانده شوند. به این ترتیب، کد فارسی پیشنهادی یک کد نسبتاً کامل بوده و به تنهایی

(بدون نیاز به قابلیت کارکرد دوزبان) در سیستمهای کامپیوتری نیز قابل استفاده است به عنوان مثال میتوان از خط عمودی و خط زیر یا بالایی حروف که علاوه بر استفاده های دیگر، در رسم جدول کاربرد دارند و یا بر انتهای باز و بسته نامه برد. (توصیه میشود که برای حفظ یکتوختی با کاربردهای لاتین، از خط تیره (ستون ۶ سطر صفر) برای رسم جدول استفاده نشود چون این علامت در جدول کد لاتین موجود نیست). یک نکته قابل توجه این است که چون علامت ستون ۲ سطر ۱۵ در فارسی به عنوان ممیز بکار می رود، علامت تقسیم فارسی نیز باید در جدول منظور شود. البته قرار دادن بعضی از علائم مخصوص لاتین در جدول ممکن است در مواردی که به علت استفاده از کد هشت بیتی این علائم نیز مستقیماً در اختیار استفاده کننده قرار دارند مناسب به نظر آید، ولی حتی در این موارد نیز دلیل خوبی برای این کار وجود دارد. ملاحظه کنید که در کارکردن با فیلدهای اطلاعاتی فارسی اغلب احتیاج به دوازده کردن اطلاعات جهت ذخیره و یا چاپ (نمایش) وجود دارد. در صورتیکه در داخل یک فیلد فارسی، علائم مخصوص با تغییر حالت (Mode) سیستم به حالت لاتین وارد شوند، این اشکال پیش می آید که حدود فیلد فارسی به صورت صحیح جهت عملیات مربوط به وارونه کردن اطلاعات مشخص نخواهند بود.

در کلیه موارد سعی شده است که علامتهای معادل و یا مشابه در جدولهای کد فارسی ولاتین در محلها یی متناظر جدول قرار گیرند. از حروف الفباء، ارقام و علائم نقطه گذاری (که با وجود صورتهای متفاوت، عیناً در محلها یی متناظر جدول هستند) گذشته، فقط تغییرات زیر نسبت به جدول کد استاندارد IRV - ISO ملاحظه میشود: علامت ریا ل به جای علامت واحد پولی، علامت تقسیم به جای علامت "&"، علامتهای گیومه باز و بسته به جای علامتهای بزرگتر و کوچکتر، علامت ضرب به جای علامت "@"، حرکتهای حروف به جای علائم ستون ۵ سطرهای ۱۱ تا ۱۴، تنوین به جای علامت ستون ۵ سطر ۱۴، خط تیره به جای



علامت ستون ۶ سطر فروبا لایحه تشدید و همزه (روی "یا" "ه") به جای علائم ستون ۲ سطرهای ۱۱ و ۱۲. یک نتیجه مهم این تناظر در مورد سیستمهای ظاهری می شود که به عنوان مثال به علت حجم کم اطلاعات خروجی فارسی، تعویض زنجیر دستگاه چاپ به خط فارسی فقط در مواقع معینی از روز صورت می گیرد. تناظر فوق به برنامهنویسان این امکان را می دهد که عمل نوشتن و قسمت عمده ای از غلط گیری برنامهای خود را بدون احتیاج به قابلیت چاپ فارسی انجام دهند چون در خروجیهای حاصله، حرفی یا عددی بودن اطلاعات، بسیاری از علائم مخصوص جدول بندیها و اطلاعات کلی دیگر قابل تشخیص هستند.

استفاده از کد استاندارد پیشنهادی

کد توسعه یافته پیشنهادی که آن را INSIC (National Standard Information Code - Iranian) می نامیم، در محیطهای هفت بیتی یا هشت بیتی قابل استفاده است. نحوه عمل مانند استاندارد ISO (مقاله ۳) است، به این ترتیب که معرفی مجموعه فارسی G1 توسط یک رشته تبدیل با گریز (Escape Sequence) صورت می گیرد. به منظور به ثبت رساندن این رشته، گریز در سطح بین المللی فرم سه علامتی زیر پیشنهاد می شود که انتخاب نهائی آن منوط به تأیید ISO است (۱۱):

ESC 2/13 4/13

پس از تعریف نمودن مجموعه G1 توسط رشته گریز فوق، در یک محیط هشت بیتی هیچ عمل اضافی دیگری لازم نیست و مجموعه های G1 و G0 به ترتیب با مقادیر صفر و یک در بیت هشتم کد قابل دسترسی هستند. ولی در مورد یک محیط هفت بیتی لازم است که انتقال وضعیت از یک مجموعه به مجموعه دیگر به وسیله علائم کنترلی، SO (Shift out) برای انتقال به G1 و SI (Shift In) برای انتقال به G0، صورت گیرد. برای توضیح بیشتر در این زمینه به مقاله ۲ مراجعه نمائید.

لازم به تأکید است که جدول داده شده فقط مربوط به کدنمایش علائم است و شکل ظاهری علائم را در دستگاههای خروجی مشخص نمی کند. به این ترتیب وجود و نوع علائم "ع" یا "ط" در جدول کد به منظور مشخص کردن اتصال یا عدم اتصال این حروف به حروف بعدی در پردازش اطلاعات لازم است و ممکن است سازندگان در عمل در دستگاه خروجی برای زیبایی خط از چهار رنوع "ع" و یا برای کم کردن تعداد علائم از یک نوع "ط" استفاده نمایند. در مقاله ۴ توصیه هایی در مورد شکل ظاهری حروف و علائم فارسی در دستگاههای خروجی ارائه شده اند. نکات ذکر شده در این مقاله جزئی از استانداردها

پیشنهاد ارائه شده
بر اساس استاندارد
ISO-IRV و خط فارسی
تعیین شده برای
توسعه آن به صورت
استاندارد، تهیه
شده است.

پیشنها دی نیستند و صرفاً جنبه توصیه و راهنما می دارند.

همچنین چون صفحه کلید در اکثر دستگاههای ورودی کامپیوتر نقش مهمی دارد، مقاله ۵ به ارائه توصیه هایی در مورد وضعیت کلیدها نسبت به یکدیگر در صفحه کلیدهای فارسی و دوزبانها (فارسی - لاتین) اختصاص دارد. نکات ذکر شده در این مقاله نیز جزو استانداردها پیشنهادی نیستند، البته به علت اهمیت که یک نوع اختی طرح صفحه کلیدها در تعلیم مسئولین و یاد کردن اطلاعات دارد، لازم است در آینده موضوع تهیه استاندارد برای صفحه کلیدها مورد بررسی قرار گیرد.

فهرست مآخذ:

- (۱) برنامه جامع انفورماتیک در سطح کلان دفتر تحقیقات عملیاتی، معاونت انفورماتیک سازمان برنامه و بودجه، بهمن ۱۳۵۶
- (۲) Mavaddat, F. and B. Parhami, "Informatics in Iran: Problems and Prospects", Proc. of the International Conf. on Computer Applications in Developing Countries, Bangkok, Thailand, Aug. 1977, pp. 124-133
- (۳) International Standards Organization, "Seven-Bit Coded Character Set for Information Processing Interchange", ISO 646, 1973.
- (۴) American National Standard Code for Information Interchange, ANSI X3.4, 1977.
- (۵) International Standards Organization, "Code Extension Techniques for Use with the ISO Seven-Bit Coded Character Set", ISO 2022, 1973.
- (۶) Acona, J.P., S.M. Garland and J.J. Tropas, "At Last: Standards for Keyboards", Datamation, Vol 17, No. 5, pp. 32-36, March 1, 1971.

(7) Parhami, B. and F. Mavaddat, "Computers and Farsi Language: A Survey of Problem Areas", Information Processing 77 (Proc. of IFIP Congress), North-Holland, Amsterdam, 1977

(8) Parhami, B., "Impact of Farsi Language on Computing in Iran", Mideast Computer, Vol 1, No. 1 pp. 6-7, Sep. 18, 1978.

(9) Hyder, S.S., "A System for Generating Urdu/Farsi/Arabic Script", Information Processing 71 (Proc. of IFIP Congress), North-Holland, Amsterdam 1972

(10) Parhami, B., "On the Use of Farsi and Arabic Languages in Computer-Based Information Systems", Proc. of Symp. on Linguistic Implications of Computer-Based Information Systems, New Delhi, Nov. 1978.

(11) International Standards Organization, "Procedure for Registration of Escape Sequences", ISO 2375, 1974.

(۱۲) حروف فارسی در ماشینهای تحریر، موسسه استاندارد و تحقیقات صنعتی ایران، استاندارد شماره ۸۲۰، آذرماه ۱۳۵۵

ضوابط و بهای درج آگهی در گزارش کامپیوتر

اندازه های مجاز (قبل از کوچک شدن):

پهنای - ۸، ۱۷، یا ۲۶ سانتیمتر
بلندی - هر مقدار تا حداکثر ۳۵ سانتیمتر

بهای درج برای یک نوبت

هر سانتیمتر بلندی در عرض یک ستون ۴۰۰ ریال
ربع ستون (۸x۸ سانتیمتر) ۲۵۰۰
نصف ستون (۱۷x۸ سانتیمتر) ۵۰۰۰
یک ستون کامل (۲۵x۸ سانتیمتر) ۱۰۰۰۰
نصف دو ستون (۱۷x۱۷ سانتیمتر) ۱۰۰۰۰
دو ستون کامل (۲۵x۱۷ سانتیمتر) ۲۰۰۰۰
نصف صفحه (۲۶x۱۷ سانتیمتر) ۱۵۰۰۰
تمام صفحه (۲۶x۲۵ سانتیمتر) ۳۰۰۰۰

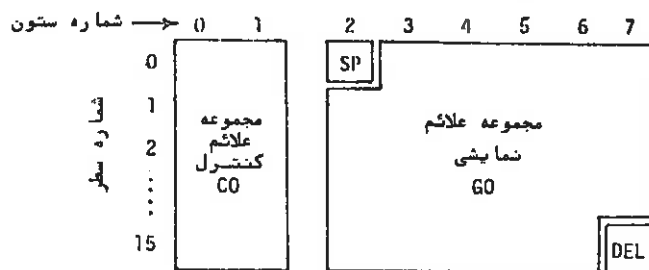
برای درج در صفحه آخر، ۲۰ درصد بهای فوق اضافه می شود.

شرایط پذیرش

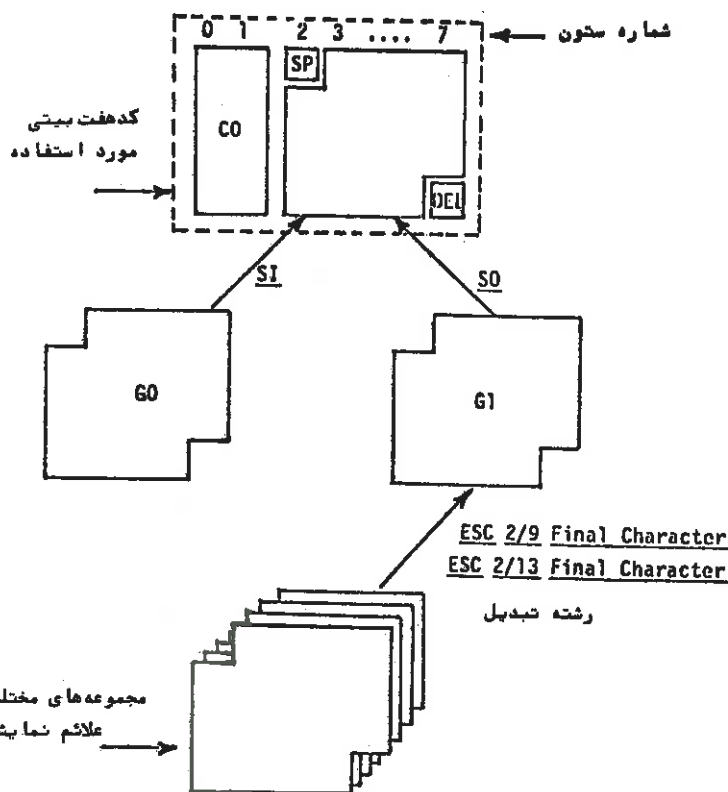
آگهی باید به صورت آماده برای چاپ حداکثر تا سومین چهارشنبه ماه قبل از انتشار به دفتر مجله برسد و محتوی آن متناسب با شأن یک مجله علمی باشد. گسسته گزارش کامپیوتر در قبول یا رد آگهیها آزاد است

رشته‌های ESC 2/9 FC و ESC 2/13 FC را رشته‌های تبدیل یا رشته‌های گریز (Escape Sequences) می‌نامیم که به منظور مشخص نمودن مجموعه G1 از بین مجموعه‌های مختلف ممکن بکار می‌روند. هر یک از علائم ستونهای سوم تا هفتم در جدول G0 می‌توانند به جای علامت نهائی FC در رشته تبدیل قرار گیرند. علائم ستون سوم برای استفاده‌های اختصاصی و علائم ستونهای چهارم تا هفتم (جمعا ۳۳ علامت) برای وضع نمودن استانداردهای توسعه یافته

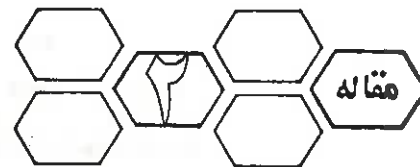
که مجموعه علائم G0 مورد استفاده قرار می‌گیرد. ESC (مخفف Escape) برای تعریف کردن علائمی که مجموعه G1 را تشکیل می‌دهد. اگرچه روشهای توسعه علائم کنترل و غیره نمودن تعریف G0 نیز در استانداردهای مورد بحث پیش بینی شده است (ماخذ از مقاله ۱) اما در اینجا به خاطر عدم احتیاج مورد بررسی قرار نمی‌گیرند. شکل ۲، روش کار با کد توسعه یافته را در یک محیط هفت بیتی نشان می‌دهد.



شکل ۱: دست‌پنویسی علائم در استاندارد ISO

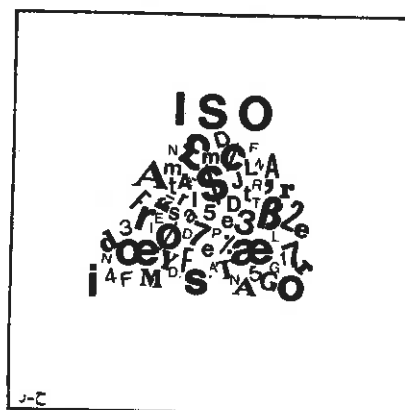


شکل ۲: روش کار با کد توسعه یافته ISO در یک محیط هفت بیتی



مختصری در باره روشن‌استاد برای توسعه کد ISO

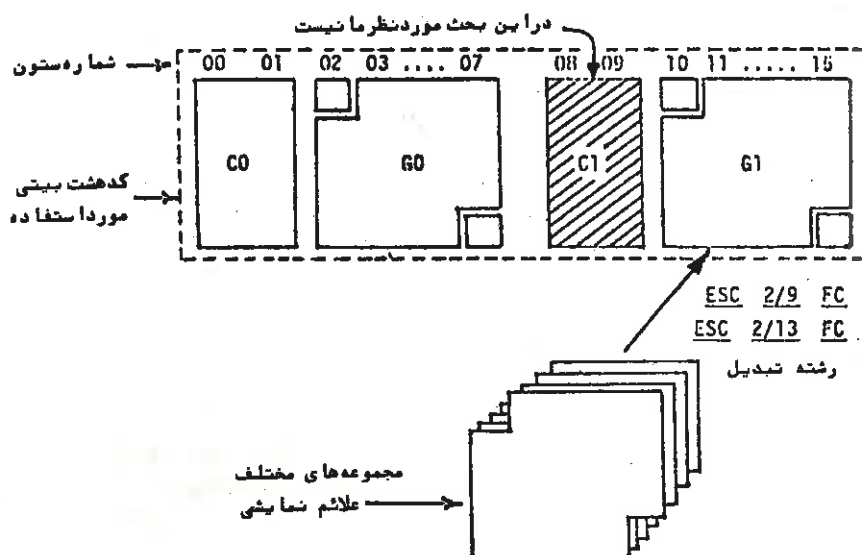
کمیته مطالعاتی استاندارد
تبادل اطلاعات در فارسی
سازمان برنامه و بودجه.



کارکرد دوزبانه / کارکرد فارسی / جدول
رمز نویسه‌ها / استاندارد سازی

کد استاندارد ISO هفت بیتی از نظر تقسیم علائم کد به دسته‌های مختلف دارای سازمانی به صورت شکل ۱ است. اضافه کردن تعداد علائم کنترل و نمایشی با بیدمطابق با قوانین استاندارد توسعه کد ISO صورت‌گیرد تا سیستم‌هایی که از کدهای توسعه یافته استفاده می‌کنند، همسازي خود را از دست ندهند. در اینجا ما فقط روش اضافه کردن تعداد علائم نمایشی را مورد بحث قرار می‌دهیم. برای توسعه کد ISO سه علامت مخصوص در بین علائم کنترل T در نظر گرفته شده‌اند. این علائم عبارتند از: SO (مخفف Shift Out) برای استفاده از مجموعه جدیدی از علائم نمایشی بنام G1 که قبلاً تعریف شده است. SI (مخفف Shift In) برای بازگشت به حالتی

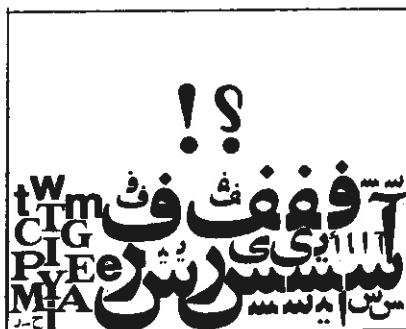
اضافه کردن تعداد علائم کنترل و نمایشی باید مطابق با قوانین استاندارد توسعه کد ISO صورت گیرد تا سیستمهایی که از کدهای توسعه یافته استفاده می کنند همسازی خود را از دست ندهند.



شکل ۲: روش کار با کد توسعه یافته ISO در یک محیط هشت بیتی

قابل استفاده هستند، با این ترتیب تنها ۱۲۶ استاندارد توسعه یافته قابل تعریف است. بنا بر این در یک محیط هشت بیتی، مجموعه G1 مورد نظر خود را توسط رشته تبدیل مربوطه تعریف کرده و آن پس با کمک علائم کنترل SO و SI از علائم موجود در G1 و G0 استفاده می کنیم.

در یک محیط هشت بیتی نحوه عمل تا حدودی ساده تر است و به صورت شکل ۳ بیان می شود. در اینجا نیز با کمک رشته های تبدیل که بر حسب قرارداد مشخص شده اند می توان مجموعه های مختلفی را به عنوان G1 تعریف نمود و مسوود استفاده را در ادکلیمه مطالبی که در مورد کد هفت بیتی ذکر شد در این مورد نیز صادق است. با این تفاوت که چون G1 و G0 هر دو در فضای کد موجودند استفاده از علائم کنترل SO و SI لازم نیست، بر حسب قرارداد اگر فقط یک مجموعه از علائم برای G1 مورد نظر باشد می توان رشته های تبدیل را نیز حذف نمود ولی این کار امکان توسعه های بعدی را از بین می برد.



در قدیم معمولاً کامپیوترها فقط فارسی بودند که تنها با علائم یک زبان (انگلیسی و عمومیتر بگوئیم، اصطلاحاً لاتین) کار کرده و فرا میسند و دستورات آنها نیز منتهی از زبان مزبور بود. امولادستورات کلیمه زبانهای برنا مهنویسی رایج همگی به زبان انگلیسی شبیه بودند، در این میان کشورهای که زبان نشان غیر لاتین بود به داشتن خروجیهای چاپ کامپیوتر به زبان خود بسنده می کردند، در آن حالت کلیمه ورودیها، پردازشها و عملیات به زبان لاتین صورت می پذیرفت و فقط بر روی کاغذ، علائم زبان مورد نظر چاپ می گردید، به این ترتیب که گوی بسا

بقیه در صفحه ۲۴

کارکرد فارسی و دو زبان بر روی کامپیوتر

حمید رضا رهبر

کارکرد دو زبان / کارکرد فارسی / جدول رمز نویسه ها / روش تک تبادی / دستگاه های ورودی - خروجی / پردازش اطلاعات

توضیح نویسنده

مقاله زیر برگزیده ای از دو مطلب جداگانه است که قبلاً برای دومنظور مستقل و برای استفاده افراد نا آشنا و یا کم آشنا با کامپیوتر نوشته شده بود و اینک به عنوان اولین مقاله از این بخش جدید نشریه اندکی تغییرات به چاپ میرسد، به همین دلیل نیز بخشی از این مطلب قبلاً درجائی دیگر به چاپ رسیده است.

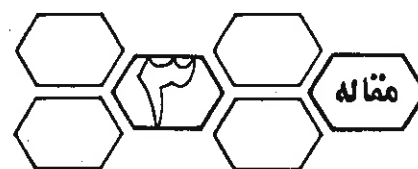


نقطه شروع

مبتدیان

یادداشت سردبیر

از این شماره، بخش جدیدی در نشریه به نام "نقطه شروع مبتدیان" ایجاد شده است که سعی می شود همه ما هه نیز با آن آشنا شویم. در این بخش ما به بررسی مفاهیم اساسی علوم کامپیوتر به زبان ساده خواهیم پرداخت به طوری که برای خوانندگان مبتدیان، آسانی و آهسته و آهسته یادگیری و آشنایی با این فیلد فراهم شود. سعی بر آن است که مطلب هر ماه این بخش در رابطه با موضوع اصلی و جهت نشریه همان ماه باشد آن چنان که به عنوان کلیدی جهت آشنایی عمومی با کلیات موضوع مقالات آن شماره به کار رود. امید است که از این طریق، مقالات هر شماره برای گروه بیشتری از خوانندگان قابل استفاده باشند.



در کد تک نمادی، هر یک از حروف الفبای خط به یک صورت ذخیره و منتقل می شود و تعیین شکل آن در دستگاه خروجی به وسیله بررسی حروف اطراف آن صورت می گیرد.

مختصری در باره کد تک نمادی برای الفبای فارسی

کمیته مطالعاتی استاندارد
تبادل اطلاعات در فارسی

سازمان برنامه و بودجه

مسئله دیگری نیز در جهت عکس مسئله فسوق موجود است به این ترتیب که ممکن است در مواردی بخوابیم حروفی را که به وسیله علامت فاصله از یکدیگر جدا شده اند به صورت متصل نمایش دهیم، مثلاً برای نوشتن کلمه "مسعود" با کمی فاصله بین حروف مجاور، لازم است به صورتی مشخص نمایش دهیم که اگرچه بعد از حرف "م" علامت فاصله آمده است منظور ما استفاده از "م" اول است و نه "م" تنها، به این منظور استفاده از علامت مخصوصی بنام اتصال مجازی (Pseudo - Connection یا باختصار PC) پیشنهاد شده است (ماخذ: مقاله ۱).

به این ترتیب می توان کلمه مورد نظر را به صورت زیر نمایش داد:

بجای نیم فقط در صورتی که این کار غیر ممکن باشد (مثلاً در مورد حرف "د" در مثال فوق) فرم جدای حرف را مورد استفاده قرار دهیم. قانون کلی فوق در مواردی که عمداً "بخواهیم" حروفی را که امکان چسبیدن به یکدیگر دارند جدا بنویسیم تولید اشکال می کند، در زیر چند نمونه از این اشکالات دیده می شود:

دانش آموز حیلگر بی ادب هم منزل کم کم نخست وزیر پس انداز خط کش البته در صورتی که در محل جدایش، علامت فاصله (SP) را بکار ببریم اشکالی موجود نخواهد بود ولی استفاده از فاصله برای جدا کردن اجزای یک کلمه مجاز نیست و باید آن را فقط برای جدا کردن کلمات مختلف در متن فارسی مورد استفاده

از آن جایی که در این
روش شکل هر حرف
بستگی به حرف بعدی
نیز دارد، تشخیص
شکل صحیح حروف با
یک حرف تاخیر صورت
می گیرد.

شخص
گلج

ح-ر

م PC SP PC س PC SP PC ع PC SP PC و SP و
البته نظریه این که چنین استفاده ای در عمل به ندرت پیش می آید، استفاده از تعداد زیادی علامت برای بیان کردن منظور فوق چندان مهم نیست.

از آن جایی که با توجه به بحث های فوق، شکل هر حرف بستگی به حرف بعدی نیز دارد، تشخیص شکل صحیح حروف با یک حرف تاخیر صورت می گیرد، در دستگاه های نمایشی می توان یک شکل موقتی برای هر حرف در نظر گرفت که با وارد شدن حرف بعدی در صورت لزوم به فرم نهایی اصلاح می شود در بعضی از موارد می توان تاخیر ممکن است از یک حرف نیز بیشتر باشد، مثلاً در صورتی که استفاده از حرکات (مطابق پیشنهاد را نه شده در مقاله ۱) مورد نظر باشد، فرم نهایی حرف "م" در عبارتهای "نام را و" و "نام را و" پس از وارد کردن دو حرف بعد (فاصله در مورد مثال اول و "ه" در مورد مثال دوم) مشخص می شود.

با توجه به مثال های فوق واضح است که تعیین شکل نهایی حروف در خروجی محتاج مقدماتی عملیات پردازش در محل تصمیم گیری است.

قرارداد،
با این جهت در عمل لازم دیده شده که علامت مخصوصی به نام فاصله مجازی (Pseudo-Space یا باختصار PS) تعریف شده و در موارد فوق مسرود استفاده قرار گیرد. نقش PS عبارت است از ایجاد جدایش بین حروف مجاور بدون این که فاصله ای بین آنها بیاندازد، به عنوان مثال کلمه "دانش آموز" به صورت زیر نشان داده می شود:

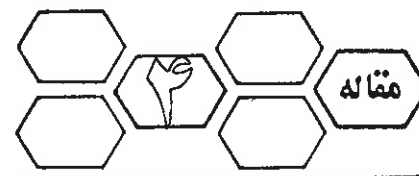
دانش PS آموز
در عمل لزوم استفاده از علامت PS کار را در کردن اطلاعات را تا حدودی مشکل می کند چون تشخیص این که این علامت در کجا مورد لزوم است احتیاج به تفکر در ادو یک عمل صرفاً مکانیکی نیست البته با جدگفت که نفس استفاده از کد تک نمادی برای تبادل اطلاعات استفاده از آن را برای دستگاه های ورودی ایجاد نمی کند، مطالعه در زمینه استانداردهای مناسب برای صفحه کلیدها (مقاله ۵) نشان داده است که بهترین است از ترکیبی نظیر صفحه کلید ماشین تحریر استفاده نمود که در آن تغییرات جزئی برای تطبیق بهتر با کار بربردهای کامپیوتری حاصل شده اند.

کارکرد فارسی / روش تک نمادی / استاندارد سازی / جدول رمز نویسی ها

در کد تک نمادی (Reduced Code Set) برای الفبای فارسی که برای اولین بار توسط هیئت (ماخذ: مقاله ۱) پیشنهاد گردید هر یک از حروف الفبای فقط به یک صورت ذخیره و منتقل می شود و تعیین شکل آن در دستگاه خروجی (مثلاً) دستگاه چاپ یا ترمینال) به وسیله بررسی حروف اطراف آن صورت می گیرد، به عنوان مثال، هرگاه در یک کلمه حروف "م" "س" "ع" "و" "د" به ترتیب ظاهر شوند، شکل حروف "س" و "ع" به صورت متصل از دو طرف در نظر گرفته می شود، چون اولی به "م" و "و" دوم به "س" و "و" می چسبند، پس حروف در دستگاه خروجی به صورت زیر ظاهر می شوند:

مسعود

قانون کلی در تعیین شکل هر حرف این است که تا حد ممکن سعی شود آن را به حروف دو طرف



نگاهی در باره نمایش علائم فارسی در دستگاههای خروجی

کمیته مطالعاتی استاندارد
تبادل اطلاعات در فارسی

سازمان برنامه و بودجه

اگرچه صورتیهای فوق تا حدودی قابل قبولند ولی با خط صحیح فارسی تطبیق نمیکنند. بنا براین توصیه میشود که در صورت امکان گونه‌های اضافی زیر نیز برای حروف فوق الذکر (هنگامیکه همراه با "دنباله‌گود" مورد استفاده قرار میگیرند) در نظر گرفته شوند تا فرم صحیح حروف بزرگ مربوطه تولید گردد:

ش ش ه

۴ - حروف "ظ" و "ط" در ماشینهای تحریر فارسی نقطه دارای یک گونه هستند، بنابراین برای نوشتن کلمه‌ای مانند "خط کش" به طوری که دو جزء کلمه به هم نچسبند، ماشین نویسنده ناگزیر است از علامت فاصله استفاده کند. ولی این کار هم از نظر خط فارسی صحیح نیست و هم در کار برگردانی به صورت متن برای تشخیص و جدا کردن کلمات تولید اشکال می‌کند. بنا براین توصیه می‌شود که از دو گونه "غیر متصل" و "متصل" این حروف به صورت زیر استفاده شود:

ط ط ظ ظ

آنها چهار گونه قابل می‌شویم). در مورد حروف چهار گونه‌ای، گونه‌ها را "تنها"، "آخر"، "اول" و "وسط" نامگذاری می‌کنیم (همزه فاقد گونه "آخر" است). اگرچه استفاده از تمام گونه‌های مختلف حروف در دستگاههای خروجی کامپیوتر امکان ایجاد خط زیبا و خوانا تری را به ما می‌دهد، در صورت دقت کافی در طرح شکل حروف، می‌توان در اکثر موارد با نصف تعداد گونه‌ها نتایج رضایت بخشی به دست آورد. در این قبیل موارد، گونه‌های "تنها" و "آخر" و همچنین گونه‌های "اول" و "وسط" در هم ادغام می‌شوند. اصطلاحاً این دو فرم را "بزرگ" و "کوچک" می‌نامیم.

نمونه‌هایی از این نوع تقلیل تعداد گونه‌ها در ماشینهای تحریر فارسی موجودند. نکات زیر در مورد شکل حروف فارسی قابل ذکرند:

۱- در مورد حروف "ت" و "ا"، اگرچه با یک گونه می‌توان نتایج قابل قبولی به دست آورد، ولی با اضافه کردن گونه‌های "ت" و "ا" می‌توان به نتایج بهتری دست یافت.

در فارسی از "خط تیره"
برای کشیدن حروف
استفاده می‌شود لذا
می‌توان آن را نیز
جزو علائم الفبایی
منظور داشت.

۵ - حروف "ع" و "غ" تنها حروفی هستند که بدون صدمه زدن به زیبایی و خوانایی خط نمی‌توان تعداد گونه‌های آنها را تقلیل داد. به همین سبب نیز ماشینهای تحریر فارسی دارای هر چهار گونه‌ای این حروف به صورتی زیر هستند:

ع ع غ غ غ

چون کد استاندارد پیشنهادی به منظور رعایت اختصار رویکنواختی فقط دو گونه از هر یک از این حروف را در بردارد و تشخیص فرم خروجی فقط با بررسی حرف قبلی در متن امکان پذیر است، می‌توان در مواردی که امکان این تمایز در دستگاه خروجی ممکن نیست، مانند بعضی از سیستمهای فعلی فقط دو گونه "تنها" و "اول" از این حروف را مورد استفاده قرار داد. در این صورت نتایجی مانند مثالهای زیر تولید خواهند شد:

مانع مغایر تیغ معادل

۶ - برای علامت ترکیبی "لا" در بسیاری از ماشینهای تحریر دو گونه وجود دارد، ولی تجربه با ماشینهای تحریر یک گونه‌ای و دستگاههای چاپ کامپیوتر از همین نوع نشان داده است که با یک گونه (به صورت "لا") می‌توان زیباترینی

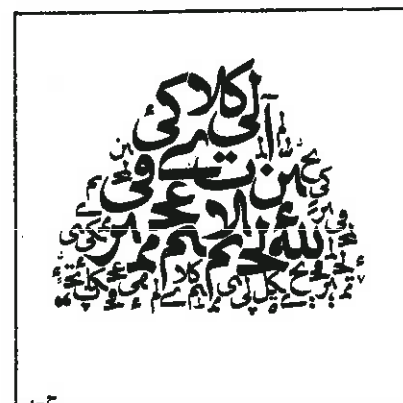
۲- زیبایی و خوانایی در خط فارسی حکم می‌کند که حروف بزرگ "ب"، "پ"، "ت"، "ث"، "ف"، "گ"، و "ک" با طول بیشتری نسبت به حروف کوچک نوشته شوند. روش متداول برای رعایت این موضوع در دستگاههای چاپ با حروف به اندازه‌ناقص است. استفاده از یک "دنباله‌تخت" همراه با حرف کوچک مربوطه است. نظریه آنچه در مورد تناظر یک به یک بین حروف در متنهای خروجی و کدهای نمایش آنها ذکر کردیم، علامت "دنباله‌تخت" به جای حروف بزرگ فوق در جدول استاندارد نگاشته شده است (ستون ۴ سطر ۱). به مثالهای زیر توجه کنید:

ب ب پ پ ت ت ث ث ف ف گ گ ک ک

این نحوه عمل مزیت تقلیل تعداد علائم لازم را نیز به همراه دارد.

۳- موارد ذکر شده در بند ۲ برای حروف بزرگ "س"، "ش"، "ص"، و "ض" نیز صادق است. در این مورد "دنباله‌گود" (ستون ۴ سطر ۲ در جدول استاندارد) همراه با حرف کوچک مربوطه مورد استفاده قرار می‌گیرد. مثالهای زیر از بعضی از سیستمهای موجود گرفته شده‌اند:

س س ش ش ض ض



ح-ر

کارکرد فارسی / جدول رمز نویسیها / دستگاههای ورودی، خروجی / استاندارد سازی

در این مورد بحث خود را به پنج قسمت تقسیم می‌کنیم: حروف الفباء، ارقام، علائم کمکی خط فارسی، علائم نقطه گذاری و علائم مخصوص.

حروف الفباء فارسی

هر یک از حروف الفباء فارسی دارای دو تنه‌ای چهار گونه مختلف است (در این میان همزه تنها حرفی است که دارای سه گونه است). حروف دو گونه‌ای عبارتند از:

ا ا د ذ ر ز و

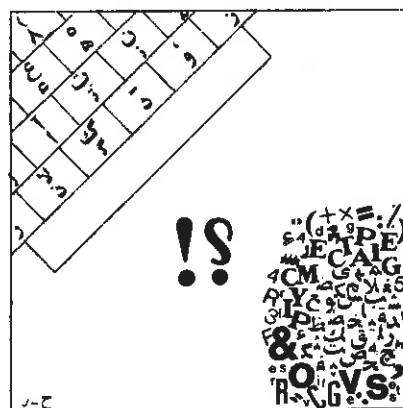
این گونه‌ها "تنها" و "آخر" می‌نامیم (در ماشینهای تحریر دستگاههای چاپ کامپیوتری فارسی موجود، حروف "ظ" و "ط" نیز جزو این دسته محسوب می‌شوند ولی نظریه‌ای همبستگی که تشخیص اتصال حرف مجاور در کار برگردانی کامپیوتری دارد، ما برای



برای صفحه کلید فارسی حداقل ۸۸ کلید لازم است.

نگاهی دوباره طرح صفحه کلیدهای فارسی و دو زبانه

کمیته مطالعاتی استاندارد
تبادل اطلاعات در فارسی
سازمان برنامه و بودجه



۲- وارد کردن هریک از حروف الفباء با فشار دادن فقط یک کلید علیرغم این موضوع که بعضی از حروف (مانند "ک") در کد استاندارد از ترکیب دو علامت حاصل میشوند (فرض بر این است که اضافه کردن دنباله های لازم به صورت خودکار انجام می شود).

۴- قرار گرفتن علائم معادل یا مشابه در صفحه کلیدهای فارسی ولاتین (بعداً در مورد آن بحث خواهد شد) در مرحله ای متناظر تا هم بتوان علائم ظاهر شده روی کلیدها را کمتر کرد و هم به خاطر سپردن محل علائم برای استفاده کنندگان از صفحه کلیدهای دوزبانه ساده تر باشد.

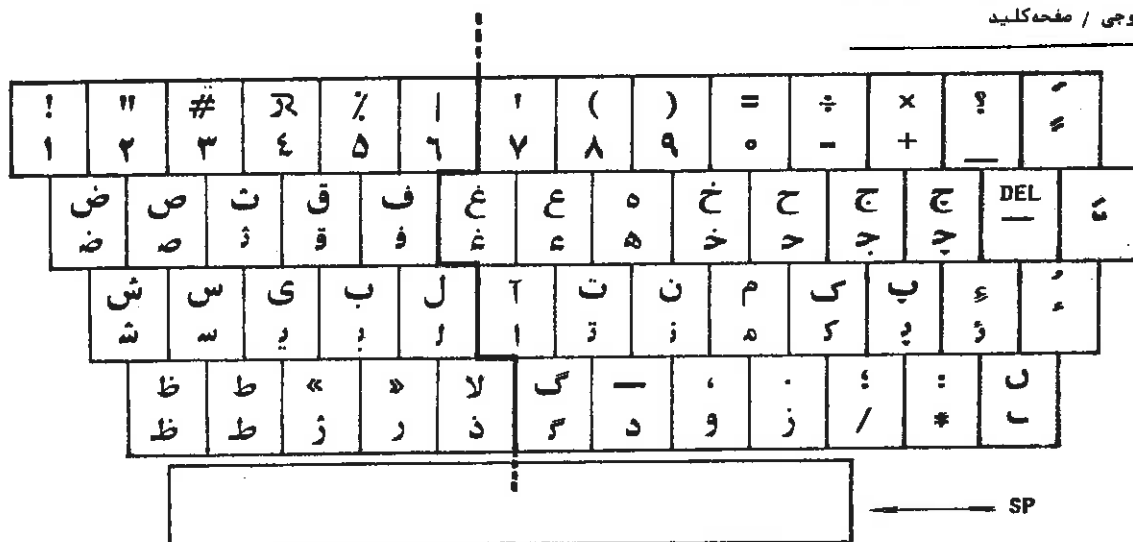
چنان که در شکل ۱ مشاهده می شود، حداقل ۴۸ کلید برای صفحه کلید فارسی لازم است (کلیدهای سفید) با اضافه کردن کلید چهل و نهم، می توان علائم ای خط بالای حروف و DEL را به علائم صفحه افزود و به این ترتیب یک صفحه کلید نسبتاً کامل برای فارسی به دست آورد، اما افزودن یک کلید دیگر در انتها، البته هریک از چهار ردیف تعداد کلیدها را به ۵۳ می رساند که برای وارد کردن کلیه علائم کد استاندارد (INSIC) کافی است. علائم دنباله تخت و دنباله گود، اگرچه در مواردی ممکن است

همان طور که برای ساده کردن امر تعلیم ماشین نویسی لازم ایجاد یک نوع اختی در طرح صفحه کلید این ماشینها احساس شود متعاقباً استانداردها برای این منظور به وجود آمد، در مورد صفحه کلید یا نه های کامپیوتری (ترمینالها) وسایردستگاههای ورودی کامپیوتر نیز همین مراحل طی گردید. متأسفانه در مورد صفحه کلیدهای فارسی به دلایل مختلف، که بحث در مورد آنها از حوصله این مختصر خارج است، هنوز هیچ نوع یکتواختی وجود ندارد. نکاتی که در این زمینه ذکر خواهد شد با بدیه عنوان یک پیشنهاد دمقدماتی تا زمان پیدایش استانداردهای لازم در این زمینه تلقی شوند.

در شکل ۱ طرح صفحه کلید فارسی پیشنهادی برای استفاده در رابطه با کد استاندارد پیشنهادی فارسی (INSIC) دیده می شود. اهم نکات رعایت شده در طرح این صفحه کلید به قرار زیرند:

- ۱- رعایت ترکیب صفحه کلید ماشینهای تحریر تا حد امکان (غیر از مواردی که این موضوع با سایر هدفهای مشروح زیر کاملاً متعارض باشد).
- ۲- دارا بودن کلیه علائم کد استاندارد پیشنهادی و فقط این علائم (به این ترتیب مثلاً "علامت" که در ماشینهای تحریر موجود است در اینجا وجود دارد).

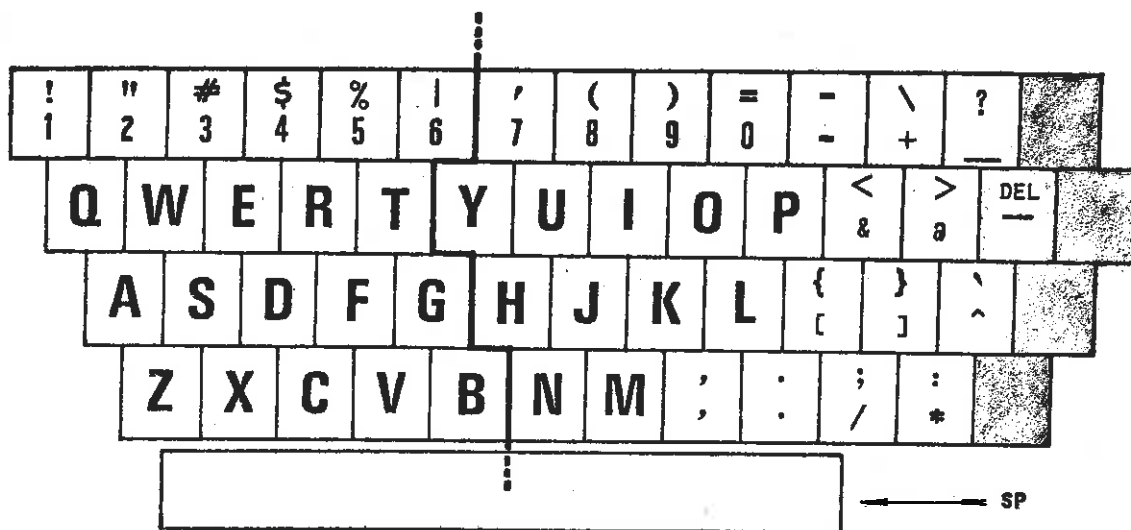
شکل ۱: طرح پیشنهادی برای صفحه کلید فارسی



همان طور که طرح متداول صفحه کلیدهای لاتین به نام ترکیب "qwerty" مشهور است طرح ارائه شده برای صفحه کلیدهای فارسی را از روی حروف ردیف پائین "زودگذر" می نامیم

کارکرد دوزبانه / کارکرد فارسی / جدول
رمز نویسه ها / استاندارد سازی / دستگاه های
ورودی - خروجی / صفحه کلید

در طرح پیشنهادی صفحه کلید فارسی و صفحه کلید دو زبانه، هماهنگی با صفحه کلید ماشینهای تحریر رعایت شده است و علائم معادل یا مشابه فارسی و لاتین نیز در محلهای متناظر قرار گرفته اند.



شکل ۲: طرح پیشنهادی برای صفحه کلید لاتین به منظور استفاده در صفحه کلیدهای دوزبانه فارسی - لاتین

فارسی و یک یا دو علامت سمت چپ (احتمالا با یک رنگ دیگر) مربوط به لاتین باشد. با توجه به محل مشترک علائم یکسان و معادل در دو طرح پیشنهادی، میتوان این علائم مشترک را فقط یکبار روی کلید نشان داد. حتی در مورد ارقام میتوان برای خلوت شدن زوی صفحه کلید، فقط ارقام لاتین را نشان داد. برای جلوگیری از هر نوع اشتباه، استفاده از رنگ سوم برای این نوع علائم مشترک مناسب است. با توجه به ۲۶ کلید از حروف فارسی که به دو گونه از یک حرف اختصاص دارند، میتوان فقط نوع بزرگ حروف را نشان داد و به استفاده کنندگان با آورش و رد کسه مانند صفحه کلید لاتین، قسمت با شین این کلیدها به حروف کوچک و قسمت با لای آنها به حروف بزرگ اختصاص دارد.



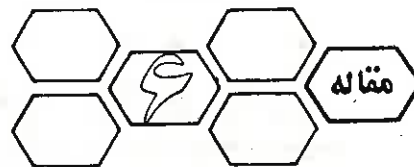
سپردن محل علائم توسط استفاده کنندگان صفحه کلیدهای دوزبانه. چنان که در طرح داده شده مشاهده میشود، چهار و هشت کلید برای این منظور کافی است و با افزودن کلید چهار و هشت میتوان مجموعه علائم کد استاندارد ISO - IRV را کامل نمود. در این طرح، مطابق معمول در صفحه کلیدهای ماشین تحریر، علائم نقطه و کاما هم در پایین و هم در بالای کلید مربوطه قرار گرفته اند و این روش در مورد علائم "منها" نیز که در کاربردهای کامپیوتری زیاده کار می رود اعمال شده است. از نظر قرار دادن شکل حروف و علائم بر روی کلیدهای صفحه کلید دوزبانه، روشهای مختلف قابل استفاده است که در اینجا بعضی از آنها را فهرست و ذکر می کنیم. روش اول این است که علائم مربوط به زبان اول یا زبان اصلی (مثلا فارسی) به صورت عادی روی کلید و علائم مربوط به زبان دوم یا زبان فرعی (مثلا لاتین) بر روی سطح جلویی کلید ظاهر شود. البته برای ایجاد تمایز بیشتر میتوان از دو رنگ مختلف برای این کار استفاده نمود. روش دوم این است که بر روی هر کلید چهار علامت (در مورد کلیدهای مربوط به حروف لاتین سه علامت) در چهار گوشه آن ظاهر شوند به طوری که مثلا دو علامت سمت راست مربوط به

کاربرد نداشته باشند، ولی چون در صفحه کلید کاما با یکدیگر همه علائم کد را تولید نمود در طرح پیشنهادی جزو کلیدهای اختیاری قرار گرفته اند. به طور کلی در اکثر موارد ۴۸ یا ۴۹ کلید برای کاربردهای معمولی کافی است.

در شکل ۲، طرح صفحه کلید لاتین پیشنهادی برای استفاده در صفحه کلیدهای دوزبانه فارسی - لاتین (همراه با طرح صفحه کلید فارسی که قبلا ارائه شد) دیده میشود. اهم نکات رعایت شده در طرح این صفحه کلید به قرار زیرند:

- ۱ - رعایت ترکیب صفحه کلید ماشینهای تحریر و صفحه کلیدهای استاندارد ASCII تا حد امکان (غیر از مواردی که این موضوع با بیزهدهای متروحه زیر کاملاً برپا شد).
- ۲ - دارا بودن کلید علائم کد استاندارد ISO - IRV (با تغییر دادن علامت واحد پولی به علامت "\$" که در ایران کاربرد نسبتاً زیادی دارد) و فقط این علائم.
- ۳ - قرار گرفتن علائم معادل یا مشابه در صفحه کلیدهای لاتین و فارسی (که قبلاً تشریح شد) در محلهای متناظر به منظور کم کردن تعداد علائم ظاهر شده روی کلیدها و سهولت بخاطر

با توجه به این که قریب هفتاد درصد و حدوداً نیمی از کامپیوترها صرف مرتب سازی و جستجوی اطلاعات می شود، ترتیب الفبایی علائم در جدول رمزبندی از اصول مهم گداستاندار است.

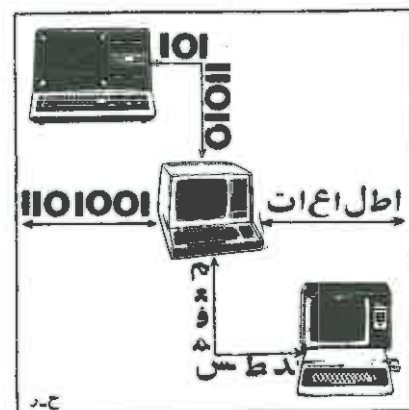


گداستاندار برای

تبادل اطلاعات

در فارسی

دکتر محمد صنعتی



کارکرد فارسی / جدول رمزنویسها / روش
تک نمادی / استنادار سازی / پردازش
اطلاعات / صفحه کلید

مقدمه

نیاز به یک گداستاندار برای تبادل اطلاعات بین سیستمهای کامپیوتری، نیاز به تکیدن دارد، چنین استناداری سالهاست که در کشورهای غربی بوجود آمده است و باعث یک همساز (Compatibility) در سطح تبادل اطلاعات بین سیستمهای مختلف کامپیوتری شده است.

برای این که سیستمهای کامپیوتری در کشور ما مورد قبول عموم قرار گیرد، نیاز به کاربردهای گوناگون آن همگان بتوانند در جهت پیشبرد کار خویش بهره مند شوند، لازم است که مردم ما با بتوانند از این سیستمها به زبان طبیعی خود استفاده کنند، این اصل ما را وادار به تدوین و ایجاد یک گداستاندار برای تبادل اطلاعات در زبان فارسی نمود که مطالعه و بحث در چگونگی تدوین آن موضوع این گزارش است.

اصول کلی تدوین استنادار

قبل از بحث و معرفی استنادار پیشنهادی، لازم است اصولی را که باید در تدوین گداستاندار برای تبادل اطلاعات در نظر گرفته شود به ترتیب اهمیت مطرح کنیم:

۱ - در تدوین گداستاندار برای تبادل اطلاعات، مهمترین اصلی که باید مورد توجه قرار گیرد، استقلال کامل گداستاندار از تکنولوژیهای خاص و سیستمهای کامپیوتری بخصوص میباشد، گذری به تاریخ کوتاه کامپیوتر و مشاهده تغییر سریع تکنولوژی آن در این مدت کم، اهمیت این اصل را آشکار میسازد، وابستگی استنادار به تکنولوژی خاص نه تنها آن را با گذشت زمان بی اهمیت میکند، بلکه باعث عدم پذیرش عمومی آن توسط سازندگان و جامعه استفاده کنندگان کامپیوتر می شود.

۲ - گداستاندار باید شامل یک مجموعه کامل باشد، این مجموعه باید کلیه علائم متداول در زبان فارسی را در بر داشته باشد، به طوری که تبادل اطلاعات و متنهای فارسی به آن گونه که مرسوم است، امکان پذیر باشد.

۳ - در تدوین گداستاندار باید به ترتیب قرار گرفتن علائم در جدول رمز، اهمیت فوق العاده ای داده شود، تا مرتب سازی اطلاعات (Sorting) به سهولت انجام پذیر باشد. حفظ ترتیب الفبایی علائم در جدول رمزبندی از اصول مهم گداستاندار میباشد. با توجه به این که قریب هفتاد درصد وقت کامپیوترها صرف مرتب سازی و جستجو (Sorting & Searching) - اطلاعات می شود، این اصل را به هیچ وجه نمیتوان نادیده گرفت، تا شایر این اصل مهم در تدوین گداستاندار، پیشنهادی به تفصیل مورد بحث قرار خواهد گرفت.

۴ - در تدوین گداستاندار باید سهولت استفاده از آن در کاربردهای نرم افزاری تا حد امکان مورد توجه قرار گیرد. چند نمونه از این موارد عبارتند از:

- حفظ ارتباط یک به یک بین علائم و گداشتهای در جدول رمز.

- دستهبندی علائم بر حسب کاربرد آنها مانند علائم الفباء، علائم عددی، علائم ریاضی و غیره.

- ترتیب علائم در جدول رمز به نحوی که تشخیص علائم هم نوع، مانند ارقام، از سایر علائم به سهولت امکان پذیر باشد.

بررسی کارهای انجام شده

در اینجا مختصراً به بررسی کارهای انجام شده که به تدوین گداستاندار در فارسی مربوط می شود می پردازیم:

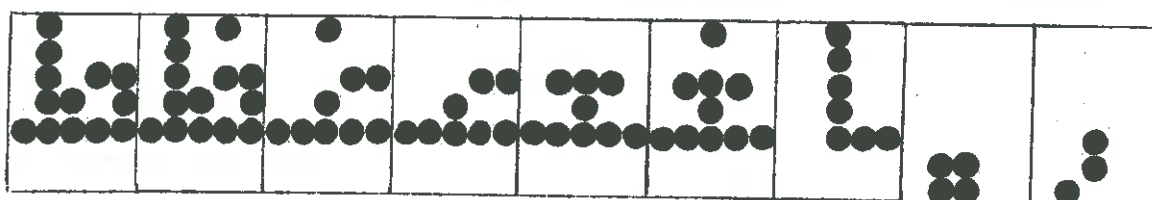
اساسیترین کوششی که در این زمینه در ایران انجام شده است مربوط به کار "کمپته مطالعاتی استنادار" تبادل اطلاعات در فارسی" است. پیشنهادیهای این کمپته در شهریور ۱۳۵۸ منتشر شده است (مقالات ۱ تا ۵). حال که چندین سال از گزارش کمپته مطالعاتی می گذرد، بررسی آن وابستگی آن را تا حدود زیادی به تکنولوژی زمان خود نمایی میسازد.

بزرگترین ایرادی که در کمپتههای کمپته مطالعاتی دیده می شود، مربوط به عدم رعایت رابطه یک به یک بین علائم و گداشتهای در جدول رمز میباشد. آنها در رابطه با دستگای چاپگر و نما بشگر زمان خوددنبالهای "تخت" و "گود" (س و ن) را در جدول رمز معرفی کرده اند، تا بتوانند علائم (ب، پ، ت، ث، س، ش، ص، ض، ک، گ) را در دستگای خروجی زیبا تر نشان دهند. معرفی این دنباله ها باعث می شود که علائم ذکر شده فوق در تبادل اطلاعات به وسیله دو کد، علامت و دنباله، نمایش داده شوند. این مسئله غیر از ازدیاد قابل توجه طول متنها باعث

اشکالات فراوانی از نقطه نظر نرم افزاری می شود که بحث آن در این گزارش مقدور نیست. اشاره به این نکته لازم است که این دنباله ها ترتیب الفبایی جدول رموز (اصل سوم) برهم میزنند و آرایش اطلاعات به جز از طریق پردازش زیاده نرم افزاری امکان پذیر نیست. برای مثال "دانش سرا" قبل از "دانشجو" ردیف می شود، که این برخلاف فرهنگ نامه های فارسی است.

با اشاره به بحث فوق یک بار دیگر اهمیت استقلال گداستاندار از تکنولوژیهای خاص را (اصل اول) تاکید می کنیم، و خاطرنشان میسازیم که مسائلی از قبیل زیبایی خط فارسی در دستگای خروجی تنها به تدوین کد استنادار برای تبادل اطلاعات نقشی بسازی کنند، باید در نظر داشت که هدف تعیین استنادار برای "تبادل" اطلاعات بین سیستمهای کامپیوتری است. نمایش زیبای متنهای فارسی در دستگای خروجی خود مسئله مهمی است که باید جداگانه مورد مطالعه عمیق قرار گیرد، که مسلماً کیفیت آن تا اندازه ای بستگی به قابلیت دستگای خروجی خواهد داشت.

با توجه به اشتراک الفبای ما با زبان عربی ابتدا به نظر می رسد که شاید بتوان گداستاندار برای زبان عربی را با تغییرات جزئی مورد



استفاده از دنباله‌های سخت و گود جهت ارائه
زیباتر حروف، باعث برهم خوردن ترتیب الفبا
حروف شده و اشکالاتی را در مرتبه سازی اطلاعات
فارسی پدید می‌آورد.

استفاده قرار داد. معروفترین استاندارد برای عربی، استاندارد 449-ASMO می‌باشد که در مجمع عمومی اکتبر ۱۹۸۲ کمیته استاندارد عرب به تصویب رسید. متأسفانه در این استاندارد پیش بینی ادغام دوزبان نشده است. با مطالعه آن به این نتیجه می‌رسیم که جا دادن ویژگیهای زبان فارسی در استاندارد 449-ASMO امکان پذیر نیست.

کدا استاندارد پیشنهادی

مقدمتاً لازم به تذکر است که در تصمیم گیریهای تدوین کدا استاندارد روشها و توصیه های کمیته بین المللی استاندارد همیشه مدنظر بوده است. بخاطر کوتاه کردن طول گزارش ازبازگویی بین توصیه ها خودداری شده است.

در تدوین کدا استاندارد نخستین کار تعیین مجموعه کامل علائم است که بتوان با آن متون فارسی را آن طور که مرسوم است نمایش داد. این مجموعه به چهار دسته علائم الفبایی، علائم عددی، علائم نقطه گذاری و علائم ویژه تقسیم شده که به ترتیب مورد بحث قرار می‌گیرند:

۱- علائم الفبایی: این دسته با پیشا مل سی و پنج حرف الفبایی معمول در فارسی باشد که عبارتند از:

آ ا ب پ ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ غ ف ق ک گ ل م ن و ه ی
هر کدام از حروف فوق، بجز (آ ا د ذ ر ز و ه ی) دارای چهار گونه "تنها"، "اول"، "وسط"، و "آخر" می‌باشند که مجموعاً ۱۶۰ گونه (نماد) را تشکیل می‌دهند.

گونه های مختلف حروف را می‌توان با استفاده از یک الگوریتم منتج از قواعد ملایی و به کمک حروف قبلی و بعدی تعیین کرد. بنا بر این برای تبادل اطلاعات احتیاجی به گنجاندن گونه های مختلف حروف در جدول رمز نیست. هم چنین، همان طوری که قبلاً نیز اشاره شد، بهتر است که کار نمایش گونه های مختلف هر حرف در دستگا ه های خروجی با توجه به قابلیت آنها انجام شود. حال این سؤال مطرح است که آیا با کد تنگ گونه ای (تنگ نمادی) می‌توان متون فارسی را به درستی مبادله کرد؟ به عبارت دیگر آیا تنها با

استفاده از الگوریتم ملایی می‌توان متون فارسی را کد با کد تنگ گونه ای نمایش داده شده اند به صورت اصلی بازسازی کرد؟ متأسفانه جواب این سؤال منفی است. برای مثال کلمات "داداش زاده"، "مومع سرا" و "لک لک" را در نظر بگیرید. این کلمات با استفاده از کد تنگ گونه ای به صورت "د ا د ا ش ز ا د ه"، "ص و م ع ه س ر ا" و "ل ک ل ک" در می‌آیند که در موقع نمایش در دستگا ه های خروجی با به کار بستن الگوریتم ملایی به صورت "داداش زاده"، "مومع سرا" و "لک لک" ظاهر می‌شوند.

با توجه به ازدیاد کلمات مرکب در زبان فارسی، اشکال فوق استفاده از کد تنگ گونه ای برای تبادل اطلاعات را غیر ممکن می‌سازد. حل این مشکل با استفاده از علامت فاصله، یا فاصله مجازی (Pseudo Space) برای جدا کردن اجزاء کلمه ممکن نیست. این راه حل نه تنها باعث ازدیاد طول متن می‌شود، بلکه نظام الفبایی کلمات را نیز به هم زده و کار مرتب سازی اطلاعات را دچار اشکال می‌کند.

برای حل این مشکل با بدقت در با ششم در صورت تمایل، در جا هایی که حرفی امکان چسبیدن به حرف بعدی را دارد، آن حرف جدا نوشته شود. برای این منظور کد دو گونه ای را معرفی می‌کنیم. در این کد برای هر حرف دو گونه اصلی و مبدل (Shifted Code) تعریف می‌شود. از کد مبدل فقط در مواقعی استفاده می‌شود که عمداً بخواهیم از چسبیدن حرفی به حرف بعدی، برخلاف الگوریتم ملایی، جلوگیری شود. بدین وسیله در دستگا ه های خروجی در بر خورد با کد مبدل یک حرف گونه مناسب آن نمایش داده می‌شود.

این راه حل تغییری در طول متن نمی‌دهد. همان طوری که نشان خواهیم داد با جا گذاری مناسب کد دو گونه ای در جدول رمز نظام الفبایی اطلاعات تا کلاً حفظ می‌گردد. با یک عمل ساده کد مبدل تبدیل به کد اصلی آن می‌شود که کار مرتب سازی اطلاعات را دچار هیچ گونه اشکالی نمی‌کند.

در مثالهای زیر کد مبدل یک حرف با خط تیره روی آن مشخص شده است. در کد دو گونه ای، کلمات "داداش زاده"، "مومع سرا" و "لک لک" به صورت

"د ا د ا ش ز ا د ه"، "ص و م ع ه س ر ا" و "ل ک ل ک" در می‌آیند. در دستگا ه های خروجی، الگوریتم ملایی با توجه به کد مبدل خواهد توانست فرم درست کلمات را بازسازی کند. با تصمیم بر انتخاب کد دو گونه ای تعداد علائم الفبایی در جدول رمز به هفتاد می‌رسد. اگرچه بعضی حروف (آ ا د ذ ر ز و) احتیاجی به کد مبدل ندارند ولی برای سهولت کار بردهای نرم افزاری محل آنها در جدول رمز محفوظ مانده است.

مسئله دیگری که در رابطه با مجموعه الفبایی مطرح است علائمی است که در روی حروف قرار می‌گیرند و ما آنها را علائم کمکی می‌نامیم. علائم کمکی معمول در فارسی عبارتند از:

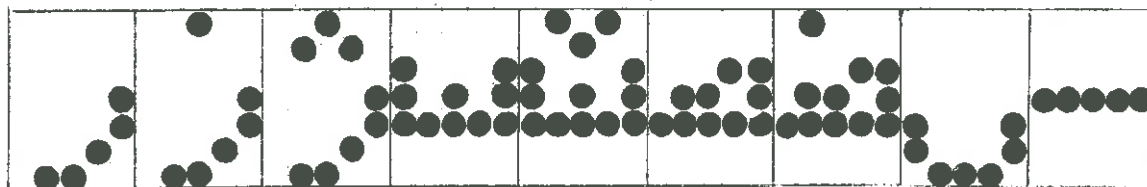
همزه: در کلماتی مانند ما، مور، مؤثر، مؤسس مأخذ.

تئوین دوزیر: در کلماتی مانند حتماً، اصلاً، سلماً

این علائم در ماشینهای تحریر فارسی مستقیماً بوده و برای مثال در هنگام ماشین کردن کلمه "مأ مور" بعد از "ما" کلید همزه بالارا می‌زنند. جا دادن این علائم کمکی در جدول رمز به طور مستقل اشکالاتی اساسی در مرتب سازی اطلاعات بوجود نمی‌آورد. برای مثال "مأ مور" (م ا م و ر) بستگی به محل "ه" در جدول رمز پیش از "ا" (م ا د ر و ا) بعد از "ما هر" (م ا ه ر) ردیف می‌شود.

مراجعه به فرهنگ نامه های فارسی نشان می‌دهد که علائم کمکی در ترتیب الفبایی کلمات نقشی بازی نمی‌کنند. به عنوان مثال کلمات "ما ما"، "مأ مور"، "مأ وا" و "ما ه" به ترتیبی که ذکر شد ردیف می‌شوند. این اشکال ما را بسیران داشت که علائم کمکی را اگر ممکن باشد به طور مستقل در جدول رمز قرار ندهیم. خوشبختانه مطالعه این مسئله نشان داد که در فارسی، همزه فقط به دو صورت "ا" و "و" و تئوین دوزیر فقط به صورت "آ" متداول است. بنا بر این تصمیم گرفته شد که این علائم کمکی همراه با حروف مربوطه شان در جدول رمز گنجانده شوند.

حال مسئله، جا گذاری صحیح این علائم در جدول علائم الفبایی بود. به طوری که نظام



زوج جدول و گونه‌های مبدل درمحل‌های فرد قرار داده شده‌اند. بنا براین تشخیص دو گونه زهم تنها با آزمای یک بیت امکان پذیر است، و با یک عمل ساده تبدیل (Shift) یا معکوس (Invert) کدمبدل یک حرف به کد ملی آن تبدیل می‌گردد.

نکاتی در مورد کد استاندارد

پیشنهادهای

فرهنگنامه‌های فارسی در ترتیب الفبایی علامت "ذ" در کلماتی مانند "مسئول" و "عائله" اختلاف نظر دارند. مثلاً فرهنگ برهان قاطع "ثین" را بعد از "آینه" می‌آورد، حال آن که در فرهنگ عمید "ثین" قبل از "آب" می‌آید. با مطالعه فرهنگنامه‌های فارسی مشاهده شد که اکثراً "ذ" را معادل "ی" در آخر الفباء قرار می‌دهند. بنا براین علامت "ذ" در جدول رمز نیز بعد از "ی" آورده شده است.

علامت ترکیبی "لا" در ماشینهای تحریر، برای تولید خط زیبای فارسی، جداگانه در روی

هفت بیتی به طور مستقل قابل استفاده باشد و هم در محیطهای هشت بیتی به عنوان توسعه استاندارد بین المللی (ISO-IRV) برای کاربردهای دوزبان بتوان از آن استفاده کرد. برای این منظور دو ستون اول (مفرویک) جدول رمز در محیطهای هفت بیتی با پدیده علامت کنترل اختتام داده شود. در محیطهای هشت بیتی به جای تکرار علامت کنترل در این دو ستون، علامت استاندارد رنگاره‌مازی (گرافیک) پیشنهاد شده است. شش ستون باقیمانده جدول ۹۶ علامت مورد بحث را در بر می‌گیرند.

شکل ۱ ساختمان جدول کد استاندارد پیشنهادی در محیط هفت بیتی و شکل ۲ ساختمان آن را در محیط هشت بیتی نشان می‌دهد.

در جایگذاری علامت الفبایی در جدول رمز سعی بر آن بوده است که این علامت‌گذاری شوند که دو گونه اصلی و مبدل یک حرف به آنسانی قابل تشخیص باشند. همچنین، با یک عمل ساده بتوان گونه مبدل را به گونه اصلی تبدیل کرد. بدین منظور گونه‌های اصلی الفباء در محیطهای

الفبایی محفوظ بماند. بدین منظور تصمیم گرفته شد که علامت "آ" در محل کدمبدل "و" و "ا" در محل کدمبدل "و" و "ز" در محل کدمبدل "و" قرار گیرد. با این جایگذاری نه تنها تعداد علامت الفبایی در جدول رمز تغییر نکرد، بلکه نظام الفبایی در موقع مرتب‌سازی اطلاعات با فرهنگنامه‌های فارسی کاملاً مطابقت می‌کند.

۲- علامت عددی: این مجموعه شامل ۳۲۱۰ می‌باشد.

۳- علامت نقطه‌گذاری: در این مجموعه هفت علامت "؛"، "؟"، "و" و فاصله قرار داده شده است.

۴- علامت ویژه: به عللی که روشن خواهد شد تعداد کل علامت را به ۹۶ محدود کردیم. بنا براین ۹ علامت باقیمانده را به علامت ویژه اختصاص دادیم:

$$\mathcal{R} = / - + () \%$$

ساختمان جدول رمز

در تعیین ساختمان جدول رمز، اصل برای این بود که کد استاندارد پیشنهادی هم در محیطهای

Column	5	4	3	2	1	0	7
Row	0	0000					گ
1	0001						ک
2	0010						ل
3	0011						پ
4	0100						م
5	0101						ف
6	0110						ن
7	0111						ع
8	1000						و
9	1001						ز
10	1010						ح
11	1011						ج
12	1100						ق
13	1101						ک
14	1110						خ
15	1111						ا

شکل ۲: جدول کد استاندارد در محیطهای هشت بیتی

Column	5	4	3	2	1	0	7
Row	0	0000					گ
1	0001						ک
2	0010						ل
3	0011						پ
4	0100						م
5	0101						ف
6	0110						ن
7	0111						ع
8	1000						و
9	1001						ز
10	1010						ح
11	1011						ج
12	1100						ق
13	1101						ک
14	1110						خ
15	1111						ا

شکل ۱: جدول کد استاندارد در محیطهای هفت بیتی

۲ - رعایت تشابه بین صفحه کلید پیشنهادی و صفحه کلیدهای مورد استفاده تا حد امکان
 ۳ - مجزا نمودن علائم الفبائی از سایر علائم
 ۴ - اختصاص یک کلید مجزا به هر علامت الفبائی، به طوری که کد مبدل به یک حرف با فشار دادن همزمان کلید مبدل و کلید آن حرف حاصل شود.

با فرضهای فوق ۳۵ کلید برای علائم الفبائی لازم است و ۲۶ علامت باقیمانده را میتوان بر روی ۱۳ کلید قرار داد. با این حساب تعداد کلیدهای لازم برای تولید تمام علائم استاندارد پیشنهادی به ۴۸ میرسد. این تعداد کلید در اکثر صفحه کلیدهای لاتین وجود دارد. بنابراین، به راحتی میتوان صفحه کلید فارسی را روی صفحه کلیدهای موجود لاتین پیدا کرد.

در تدوین استاندارد برای تبادل اطلاعات مهمترین اصل، استقلال کامل کد استاندارد از تکنولوژیهای خاص و سیستم کامپیوتری مخصوص می باشد.

!	"	،	٪	٪	:	۰	؟	(	)	-	x	۳
۱	۲	۳	۴	۵	۶	۷	۸	۹	۰	+	/	=
ص	ص	ن	ق	ف	غ	ع	ه	خ	ح	ج	چ	
ا	پ	ک	م	ن	ت	ا	ل	ب	ی	س	ش	
ء	ز	و	و	ک	د	د	ر	ز	ء	ط	ظ	

شکل ۳: طرح صفحه کلید فارسی

یک کلید قرار داده شده است. اختصاص یک کد مستقل به این علامت ترکیبی اشکالاتی در امر مرتب سازی اطلاعات ایجاد میکند. بدین سبب باید به جای آن از دو حرف "ل" و "۳" استفاده شود. لازم به تذکر است که قرار دادن علامت ترکیبی "لا" در جدول رمبزه منظور مراعات اصل سوم میباشد. در دستگاههای خروجی شکل ترکیبی آن، به هر نحوی که دلپذیر باشد، قابل نمایش است.

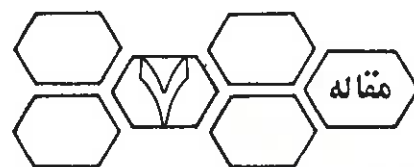
صفحه کلید استاندارد

برای سهولت در آموزش و اردک کردن اطلاعات به کامپیوتر، به یک صفحه کلید یکتا احتیاج است. در شکل ۳ طرح صفحه کلید یکتا احتیاج است. در شکل ۳ طرح صفحه کلید پیشنهادی در رابطه با کد استاندارد برای تبادل اطلاعات عرضه شده است. در طرح این صفحه کلید، تا زیر مورد توجه بوده است:

۱ - رعایت ترتیب صفحه کلید ما شینهای تحریر تا حد امکان.



بدون شرح!



مقاله

مردان غریب کارت

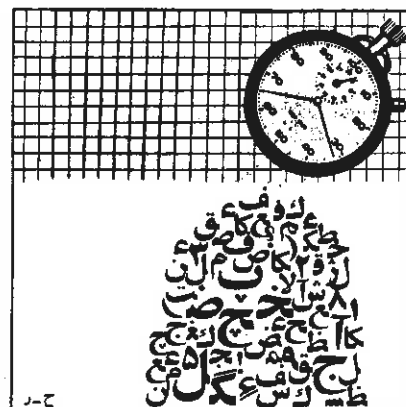
تعیین کنند مهم در تدوین استاندارد رمز تبادل اطلاعات در فارسی

قاسم جابری پور
دانشگاه شهید بهشتی

عمومیت و گستردگی برخورد را ست که رعایت کامل آن در ذخیره، پردازش و بازیابی اطلاعات فارسی در کامپیوتر چنانچه نا پذیر است، لهذا تخصیص رمز جداگانه برای حداقل دو شکل بزرگ و کوچک حروف برای امکان بازیابی «عین» آن چه ذخیره شده است «ضروری است (ماخذ ۲). از طرف دیگر در مقایسه دورشته از الفبا اگر دو شکل از یک حرف هم موضع باشند با بستی مساوی تلقی گردند در حالی که رمز آن دو متفاوت است. اگر رابطه یکنواختی بین رمزها شکل مختلف یک حرف وجود داشته باشد تشخیص همنا می بودن نویسه های دورمز مورد مقایسه از طریق تشخیص وجودی عدم وجود رابطه یکنواختی بین دورمز میسر می گردد. این رابطه یکنواختی با بستی به نحوی تعیین گردد که پیاده سازی الگوریتم مقایسه رمز نویسه و تشخیص این رابطه، در سطوح مختلف برنا مه سازی به زبان سطح بالا، سمبلی و مقایسه سخت افزاری، به ساده ترین صورت ممکن انجام گیرد.

(ماخذ ۱) و برنا مه های سیستم های انواع بسته های نرم افزاری به منظور تسهیل و گسترش استفاده از کامپیوتر برای انجام موزاداری، مالی، خدماتی، علمی، فرهنگی، آموزشی و اطلاعاتی در ایران ضروری می نماید. با توجه به این که بیشتر زمان پردازش در اغلب برنا مه های سیستم و کاربرد صرفاً مورد مرتب سازی و جستجو (Sort & Search) می شود استناد در رمز تبادل اطلاعات با بستی به گونه ای باشد که پیاده سازی الگوریتم های مرتب سازی و جستجو را به صورت بسیار آسان و امکان سازد و از آنجا که هسته اصلی این الگوریتم ها، الگوریتم مقایسه دورشته نویسه است (ماخذ ۲) رمز نویسه ها با بستی از نظم خاصی برخوردار باشند تا موجب حداقل رساندن پردازش لازم برای مقایسه دورشته نویسه باشد (ماخذ ۲) عدم وجود نظم یکنواختی در پیشنهادهای مطرح شده تا کنون را بررسی می کنند. همچنین با توجه به امکان ایجاد تکنولوژی داخلی در زمینه سخت افزار و گسترش تولید

در مقایسه دورشته از
الفبا اگر دو شکل از
یک حرف هم موضع
باشند، با بستی مساوی
تلقی گردند هر چند که
رمز آن دو متفاوت
باشد.



رمز حروف کوچک و بزرگ با اختلاف واحد

ماخذ ۲ و ۳ با اختتام رمزهای زوج به حروف بزرگ و رمزهای فرد به حروف کوچک، به طوری که اختلاف رمزهای دو شکل از یک حرف برابر یک (رمز کوچکتر برای شکل بزرگ و رمز بزرگتر برای شکل کوچک) باشد، رابطه یکنواختی را بین آنها بوجود آورده اند که از دو قسمت تشکیل شده است:

- الف - "اختلاف واحد" دورمز
ب - زوج بودن رمز کوچکتر

پیاده سازی الگوریتم تشخیص این رابطه دو قسمتی در سطوح مختلف، کارآیی خیلی خوبی را در مقایسه با انجام عمل ساده تقریبی برای یافتن نتیجه مقایسه دو نویسه در رمزهای تک نمادی بدست نمی دهد (شکل ۳).
شکل ۱، جدول رمزها "اختلاف واحد" را - نشان می دهد.
جدول شکل ۲، اختتام رمزهای ۶۴ تا ۹۵ به حروف کوچک و رمزهای ۹۶ تا ۱۲۷ به حروف

بزرگ و نرم افزاری، استناد در رمز تبادل اطلاعات فارسی با بستی به نحوی تدوین شود که طراحی سخت افزاری و نرم افزاری خاص کار بردهای فارسی (البته به صورت همسان برای کاربردهای لاتین) را مشکل و غیرکسار (Inefficient) نشاند.

اصول لازم دیگری (ماخذ ۳) معمولاً در تدوین چنین استنادها رعایتی رعایت شده و می شوند (مثل اختتام رمز متوالی به نویسه های هم سنخ از قبیل ارقام، علائم معنایی، علائم نقطه گذاری، و الفبا) که از تکرار آنها صرف نظر می شود و تا کیدای ملی در این مقاله بر نحوه اختتام رمزها به نویسه های الفبائی و اشرف آن بر روی الگوریتم های نرم افزاری و سخت افزاری مقایسه دورشته از الفبا خواهد بود.

اشکال مختلف یک حرف الفبا و ارتباط رمز آنها با هم

یک حرف الفبا در اغلب زبانهای طبیعی، در کلمات و متون مختلف به اشکال گوناگون ظاهر می شود. در زبان فارسی این امر از چنان

کارکرد فارسی / جدول رمز نویسه ها / پردازش اطلاعات / استناد از دساز / سخت افزار / طرح بدایات منطقی

خلاصه

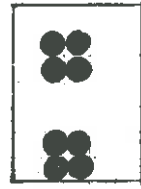
تا کنون در تدوین استناد در رمز تبادل اطلاعات در فارسی به مسأله پردازش اطلاعات فارسی توجه کافی مبذول نشده و هدف عمده (حداقل در عمل) امکان ایجاد خروجی خوانا و زیبا تر و سهولت وارد کردن اطلاعات بوده است اثر مستقیم نحوه تخصیص رمزها به نویسه های الفبائی در سرعت پردازش اطلاعات فارسی به عنوان راه نمایی خوبی در تدوین استنادها و در پیشنها جدیدی در این زمینه تلقی می شود. این بررسی با در نظر گرفتن ملاحظات نرم افزاری و سخت افزاری، همساز با استنادها و الفبائی بین المللی، و امکان بهره گیری از بسته های پیش نوشته شده خارجی صورت پذیرفته و طی آن راه حل های معمول دیگری برای وارد کردن آنها در اطلاعات و تا مین خروجی زیبا تر مطرح می گردند.

مقدمه

هر چند تعریف یک زبان برنا مه سازی همه منظوره فارسی ممکن است ضروری نباشد لکن فارسی کردن زبانهای برنا مه سازی بر کار برد



روشی "اختلاف ۳۲" در مقایسه با روشی "اختلاف واحد" کارایی بیشتری را نشان می دهد، ضمن آن که این روش بنا است بر اساس استانداردهای بین المللی همسازي دارد زیرا اختلاف میان رمزحروف کوچک و بزرگ لاتین در این استاندارد دهانویز ثابت و برابر ۳۲ می باشد.



در توضیح الگوریتم ۲ می توان گفت که اگر فقط یکی از دورمز مورد مقایسه در محدوده حروف بزرگ باشد دستور $R = R - 32$ با عمل می شود که رمز شکل کوچک آن حرف در مقایسه موثر شود و اگر فقط یکی از دورمز در محدوده حروف کوچک باشد دستور $R = R + 32$ رمز شکل بزرگ آن حرف را در مقایسه دخالت می دهد و بدین ترتیب شبیه ذکر شده در بخش قبل به سادگی و با کمترین هزینه برطرف می گردد.

ب- مقایسه معادل دوالگوریتم به زبان اسمبلی یا ریزدستورالعمل (microinstruction)

با توجه به این که انجام عملیات odd و even در زبان سطح پایین بسیار ساده و به صورت یک shift به راست و بررسی سمت راستترین بیت امکان پذیر است، الگوریتم ۲ در این سطح علیرغم کارایی بهتر، برتری قابل ذکر در مقایسه با قسمتهای "الف" و "ج" را نشان نمیدهد.

اختلاف واحد را دلکن بررسی دقیق الگوریتم های مقایسه دورشته نویسه مبتنی بر این دوروش کا رآیی بهتر و روش با اختلاف ۳۲ را آشکار میازد.

الف- الگوریتمهای مقایسه دورشته نویسه در زبان سطح بالا

شکل ۲ الف، الگوریتم مقایسه دورشته نویسه را با استفاده از روش تخصیص با اختلاف واحد (الگوریتم ۱) و شکل ۳ - ب الگوریتم مقایسه دورشته نویسه مبتنی بر جدول شکل ۲ (الگوریتم ۲) را نشان می دهد. برای مقایسه دوالگوریتم در حقیقت با یستی، انجام عملیات odd یا even در الگوریتم ۱ را به ترتیب با عمل مقایسه $>$ و $<$ یا $+$ و $-$ الگوریتم ۲ مقایسه نمود زیرا بقیه عملیات به تساوی در هر دو الگوریتم موجودند و لذا بدیهی است که الگوریتم ۲ کا رآ تر خواهد بود.

بزرگ را در یک محیط ۷ بیتی نشان می دهد. اختلاف رمز و شکل از یک حرف الفبا، ثابت و برابر ۳۲ است. تشخیص این رابطه ساده با انجام یک عمل تفریق میسر می شود. در یک نگاه ساده به این جدول، کوچکتر بودن رمزحروف کوچک از رمزحروف بزرگی که در ترتیب الفبائی از آنها کوچکترند، آشکارسال عمده ای به نظر می رسد (مثلا رمز "م" که برابر ۸ است از رمز "ش" (برابر ۱۱) کوچکتر است در حالی که در ترتیب الفبا "م" بعد از "ش" قرار می گیرد). این شبیه در الگوریتم مقایسه دورشته نویسه مبتنی بر این جدول با ظرافت برطرف شده است (شکل ۳).

مقایسه دوروش "اختلاف واحد" و "اختلاف ۳۲ در سطوح مختلف

روش تخصیص رمزهای حروف بزرگ و کوچک با اختلاف ۳۲ در ظاهرا کارایی کمتری از روش با

Column	0	1	2	3	4	5	6	7
Bit Pattern	b7 0	b6 0	b5 0	b4 0	b3 1	b2 1	b1 1	b0 1
Row	b4 b3 b2 b1							
0	0 0 0 0							
1	0 0 0 1							
2	0 0 1 0							
3	0 0 1 1							
4	0 1 0 0							
5	0 1 0 1							
6	0 1 1 0							
7	0 1 1 1							
8	1 0 0 0							
9	1 0 0 1							
10	1 0 1 0							
11	1 0 1 1							
12	1 1 0 0							
13	1 1 0 1							
14	1 1 1 0							
15	1 1 1 1							

شکل ۲: جدول تخصیص رمز نویسه های حروف فارسی با اختلاف ۳۲

Column	0	1	2	3	4	5	6	7
Bit Pattern	b7 0	b6 0	b5 0	b4 0	b3 1	b2 1	b1 1	b0 1
Row	b4 b3 b2 b1							
0	0 0 0 0							
1	0 0 0 1							
2	0 0 1 0							
3	0 0 1 1							
4	0 1 0 0							
5	0 1 0 1							
6	0 1 1 0							
7	0 1 1 1							
8	1 0 0 0							
9	1 0 0 1							
10	1 0 1 0							
11	1 0 1 1							
12	1 1 0 0							
13	1 1 0 1							
14	1 1 1 0							
15	1 1 1 1							

شکل ۱: جدول تخصیص رمز نویسه های حروف فارسی با اختلاف واحد

مقایسه کننده ۷ بیتى معمولی از دو مقایسه کننده ۴ بیتى معمولی مطابق شکل ۴ - الف به دست می آید. قسمت ب شکل ۴ نشان می دهد که با قراردادن یک مدار واسط ساده بیسند دو مقایسه کننده ۴ بیتى می توان آن را به مقایسه کننده دو منظوره عددی و حرفی تبدیل نمود که با بیت c کنترل می شود، روابط زیر مدار واسط را تعریف می کنند (برای جزئیات بیشتر به ضمیمه مراجعه شود). این روابط بر اساس نظم حاکم در استاندارد "اختلاف ۳۲" به دست آمده اند.

$$e = E \overline{C} x_6 y_6 + (x_4 \oplus y_4) C x_6 y_6$$

$$g = G \overline{C} x_6 y_6 + (x_4 \overline{y_4}) C x_6 y_6$$

به این ترتیب استفاده از مدار "ب" به جای مدار "الف" در یک برد زنده باعث می شود که دو دستورالعمل زبان ماشین برای مقایسه عددی و مقایسه حرفی که اختلاف عمل آنها فقط در تعیین بیت کنترل C (C=1 برای مقایسه حرفی و C=0 برای مقایسه عددی) می باشد طراحی نمود. البته مدار در هر دو محیط ۷ بیتى قابل استفاده است.

پایاده سازی الگوریتم ۱ در سخت افزار با مقایسه رمز و حرف از استاندارد "اختلاف واحد" به وسیله یک مدار سخت افزار به سادگی مدار فوق امکان پذیر نیست زیرا تشخیص تعلق دو رمز به دو شکل از یک حروف (و در نتیجه مساوی تلقی کردن دو رمز) در این استاندارد مستلزم به دست آوردن تفاضل دو رمز و بررسی قدر مطلق آن (که آیا برابر یا هست یا نیست) می باشد.

علاوه بر این با بستن زوج بودن رمز کوچکتر از نیز بررسی نمود، بدیهی است که ارتباط دادن تمامی این کارها به یک دستورالعمل زبان ماشین از نقطه نظر عملی در طراحی برد زنده ها مردود است و لذا برتری فوق العاده روش "اختلاف ۳۲" نسبت به روش "اختلاف واحد" از این نقطه نظر کاملاً آشکار می گردد.

همسازی با استانداردهای بین المللی

و امکان بهره گیری از بسته های پیش

نوشته شده خارجی

در استانداردهای بین المللی تبادل اطلاعات، حروف کوچک و بزرگ لاتین طوری در جدول رمز قرار گرفته اند که اختلاف رمز و شکل از یک حرف ثابت و برابر ۳۲ باشد. لذا همسازی روش "اختلاف ۳۲" (شکل ۲) با استاندارد بین المللی روشن است، در صورتی که روش با "اختلاف واحد" از این نقطه نظر کاملاً با استاندارد بین المللی همسازی ندارد. در بقیه موارد هر دو روش به یک اندازه با استانداردهای بین المللی همسازی دارند به خصوص همان طور که جدول شکل ۲ نشان می دهد، محل علائم کنترل و ارقام و علائم مخصوص با استاندارد بین المللی کاملاً یکسان است. این همسازی باعث می شود

```
R := 0; I := 1;
in: if I >= 31 then go out fi;
    if S1[I] ≠ S2[I] then go done fi;
    I := I + 1; go in;
done: R := S1[I] - S2[I];
    if S1[I] > 0 then
        if R = 1 then if odd(S1[I]) then R := 0 fi
        elsif R = -1 then if even(S1[I]) then R := 0 fi
        fi fi;
out:
```

الف - الگوریتم (۱) - مقایسه بر اساس روش اختلاف واحد

```
R := 0; I := 1;
in: if I >= 31 then go out fi;
    if S1[I] ≠ S2[I] then go done fi;
    I := I + 1; go in;
done: R := S1[I] - S2[I];
    if S1[I] >= 96 then if S2[I] < 96 then R := R - 32 fi
    elsif S1[I] >= 64 then if S2[I] >= 96 then R := R + 32 fi
    fi fi;
out:
```

ب - الگوریتم (۲) - مقایسه بر اساس اختلاف ۳۲

شکل ۳: الگوریتمهای محاسبه R، نتیجه مقایسه دو رشته نویسه S2, S1 به طول 31 به طوری که:

R = 0 iff S1 = S2
R > 0 iff S1 > S2
R < 0 iff S1 < S2

میان رمزهای اشکال کوچک و بزرگ و نیز در میان می باشد رابطه یکنواختی در میان آنها ثابت می باشد و این عمل پردازشی بدون بررسی رابطه صورت کار انجام داد.

ج - مقایسه دو الگوریتم از نظر پیاده سازی

سخت افزاری

هر چند پیاده سازی الگوریتم مقایسه دو نویسه به صورت سخت افزاری در مدارهای یک برد زنده، حداقل در آینده نزدیک به نظر می رسد و از نظر بعضی افراد ممکن است توجیه پذیر نباشد ولی امکان ساخت مدارات سخت افزاری در کاربردهای رقمی، دیگر دور از ذهن نیست لذا در تدوین استاندارد با بستن با آینده نگری و وسعت نظراتین ساله را داخل داد.

شکل ۴ تبدیل یک مقایسه کننده ۷ بیتى معمولی را (که خروجی آن نتیجه مقایسه دو عدد دودویی ۷ بیتى است به طوری که $E \oplus x = y$ و $G \oplus x > y$ و $L \oplus x < y$) به مقایسه کننده ۷ بیتى دو منظوره عددی و حرفی نشان می دهد.



به ا و ا هر حرف فارسی
ذخیره دو شکل کوچک و
بزرگ آن به عنوان
داخل و اجتناب
منه پذیر است

تشدید (۳) و خط تیره (۴) اختصار داد.
الف مقصوره (۱)، جزم (۲) و حرکات (۳)
و دیگر علائمی که ممکن است در کلمات ظاهر شوند
ولی در مقایسه دورشته نویسه دخال نکند را
می توان در محیط های ۸ بیتی به جدول اضافه
نمود. ماخذ ۲ برای کشف و حذف این گونه
نویسه ها راه حل ویژه ای ارائه داده است.

د - استفاده از محل حروف کوچک معادل

حروف د، ذ، ز، ژ

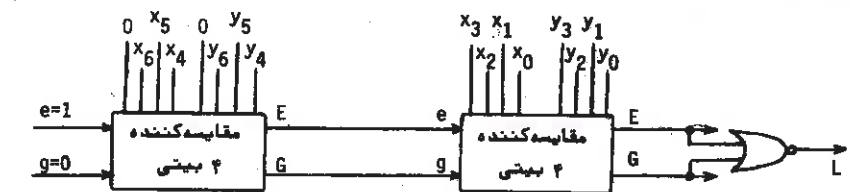
رمزهای ۷۳ تا ۷۷ به ترتیب ۲۲ واحد از
رمزهای مربوط به حروف د، ذ، ز، ژ کمتر
هستند که برای حفظ رابطه یکنواخت بین رمز
حروف بزرگ و کوچک با بیتی به شکل دیگری از
آنها (و یا دوباره به خود آنها) اختصار داده
شوند. می توان از این رمزها برای شکل چسبان
این حروف استفاده نمود.

اشکال دیگر حروف مورد استفاده در وارد کردن اطلاعات و تاسمین خروجی

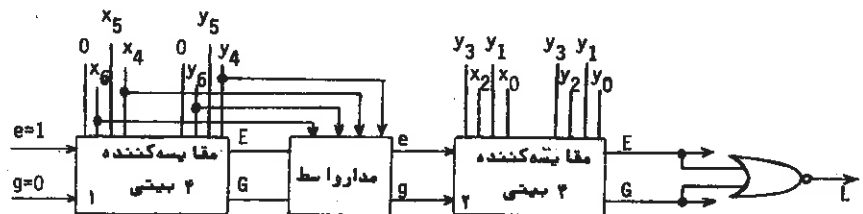
زیبا تر

شکل بعضی از حروف اگر مثلا در وسط کلمه
قرار گیرند با هیچ یک از دو شکل کوچک و بزرگ
آنها مطابق جدول شکل ۲ تطبیق نمی کند مثل
چهار تنوع عین (ع، ه، ج)، خوشبختانه نوع
شکل حرف از محل قرار گرفتن آن در کلمه قابل
تشخیص است و نیاز به تعیین رمز جداگانه برای
آن نیست (تعمیم این موضوع استفاده از رمز
تک نمادی را پیشنهاد می کند، لکن ماخذ نشان
می دهد که رمز تک نمادی اشکالات جدیدی را به
وجود می آورد).

بنابر این می توان مثلا علاوه بر اشکال ع و م
از حرف عین، اشکال د، ج را نیز بر روی صفحه
کلید جایگذاری نمود منتهی رمز (ع و ج) و رمز
(ع و د) را مساوی فرض نمود، بدین ترتیب در
هنگام وارد کردن اطلاعات به سادگی می توان
ع یا د را بر حسب این که وارد کننده اطلاعات
کلید کد امیک را فشار داده یا شد بر روی صفحه
نمایش تلویزیون دیدولی برای هر دو یک رمز
ذخیره می شود. درستگای این خروجی هوشمند
(intelligent) که از یک ریزپردازنده استفاده
می کنند به سادگی می توان یک پیش پردازشگر
تعمیه کرد تا رمز یکسان "ع" یا "د" را به علائم
الکتریکی مختلف مخصوص ع یا د (که برای
دستگاه ها پ کنده یا نمایش دهنده قابل درک
است) بر حسب نویسه های قبل و بعد آن تبدیل
نمود. بنا بر این با جدا کردن مساله پردازش از
مساله تاسمین ورودی - خروجی زیبا می توان
استاندارد مناسب برای پردازش حدا قس را



الف - مقایسه معمولی دو عدد ۲ بیتی $x = x_6 x_5 x_4 x_3 x_2 x_1 x_0$ $y = y_6 y_5 y_4 y_3 y_2 y_1 y_0$



ب - مقایسه کننده دو منظره عددی (C=0) و حرفی (C=1) با مدار واسط تعریف شده با روابط:

$$e = \overline{C} x_6 y_6 E + (x_4 \oplus y_4) C x_6 y_6$$

$$g = \overline{C} x_6 y_6 G + (x_4 \oplus y_4) C x_6 y_6$$

شکل ۴: مقایسه کننده معمولی و دو منظره

که الگوریتمهای مرتب کردن که به صورت
بسته های پیش نوشته شده در زبانهای مختلف
برنامه سازی وجود دارند و مساله کاربرد مغلط
حروف کوچک و بزرگ را حل کرده اند، عینا برای
مرتب کردن اطلاعات فارسی مورد استفاده
قرار گیرند.

خصوصیات دیگر استاندارد مبتنی بر

روش ۳ اختلاف ۳۲

با دقت در جدول شکل ۲ خصوصیات ذیل به
چشم می خورند:

الف - رمز همزه

در مورد همزه فرضیات زیر در نظر گرفته شده
است:

- همزه وسط کلمه (ث) قبل از کلمه
الفبا در نظر گرفته شده است لذا در ترتیب
معودی کلمات، کلمه "نون" قبل از "ناو"
قرار می گیرد.
- همزه وسط کلمه (ث) ممکن است به جای
ی بکار رود مانند آیین و آیین.
- الف همزه (ا) نیز قبل از کلمه الفبا
قرار می گیرد لذا در ترتیب معودی کلمات "یا س"
قبل از "یا س" خواهد بود.
- واو همزه (و) با واو (و) مساوی تلقی
می گردد، لذا در ترتیب کلمات، "مومن" یا "موم"
نزدیک هم خواهند بود.
- همزه آخر (ه) قبل از کلمه الفبا (به جز
حرف "ث" در نظر گرفته شده است، این حرف در

مقایسه دورشته نویسه مشخص لغات فارسی
هیچگاه با همزه وسط (ث) مورد مقایسه قرار
نمی گیرد زیرا اولی غیر چسبان و دومی چسبان
است و لذا نتیجه مقایسه دولفت با مقایسه
حروف قبل از همزه ها که قطعاً با هم مساوی نیستند
مشخص می گردد.

اختلاف ثابت ۲۲ همزه وسط و "ب" به
برنامه سازان می دهد که در صورت نیاز با
تبدیل اعداد ۶۳ و ۹۵ در خط آخر الگوریتم ۲
به اعداد ۶۳ و ۹۵، آنها را نیز مساوی تلقی
نماید.

ب - اشکال دیگر الف

دو شکل دیگر الف یعنی "ث" و "ا" در فارسی
به کار می روند محل "ث" قبل از تمام حروف و حتی
قبل از همزه در نظر گرفته شده است و چون درستون
۲ قرار دارد، امکان اشتباه آن با الف (ا) وجود
ندارد، به عبارت دیگر اگر کلمه "ث مد" یا "ث"
ذخیره شده یا شبیه صورت "ث مد" (الف) قابل
با زبانی نخواهد بود.

ج - تشدید، حرکات و خط تیره

دورمزا قیما نده بین محدوده ارتسام و
الفبا، یعنی رمزهای ۵۸ و ۵۹ را می توان به

باجدا کردن مساله پردازش از مساله خامی و ورودی - خروجی زیاده‌ای توان استنتاج دارد مناسب برای پردازش حداقل رابطه دست آورد در عین حال با استفاده از وسایل ورودی - خروجی هوشمند وارد کردن آسان اطلاعات و خروجی را تضمین نمود.

بهدست آورد در عین حال با استفاده از وسایل ورودی - خروجی هوشمند که گسترش یکه تا زانه‌ای پیدا کرده‌اند، وارد کردن آسان اطلاعات و خروجی زیاده‌ای تا می‌توان نمود و به عبارت دیگر رمزها را در دیکری برای وسایل خروجی در نظر گرفت که این که با زندگی این وسایل لزوم این امر را حاشی نمود و بر آن تاکید دارند (مأخذ ۴).

نتیجه

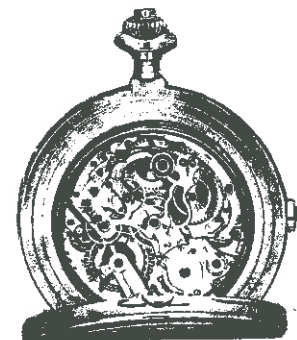
تعیین رمز نویسه‌ها، در زمان پردازش اغلب برنا مه‌ای سیستم و کاربرد اثر مستقیم دارد در تدوین استاندارد رمز تبادل اطلاعات در فارسی با بستی بیشترین توجه به تعیین رمز نویسه‌ها به نحوی که مقایسه دو حرف را از نظر ترتیب آنها در الفبا به ساده‌ترین صورت (تفاضل دورمز) ممکن سازد معطوف گردد. پیروز تدوین و تثبیت چنین استاندارد از زندگی وسایل ورودی - خروجی هوشمند به راحتی می‌توانند محصولات خود را با این استاندارد تطبیق داده و شکل حروف را کاملاً به سلیقه مشتری (سفارش دهنده) تولید نمایند.

قردانی: نگارنده بدین وسیله از آقای جواد اسماعیلی به خاطر بخشش سودمندشان و از مسئولین مرکز تحقیقات مخابرات ایران به خاطر پشتیبانی از این تحقیق صمیمانه قدردانی می‌نماید.

مآخذ:

- ۱ - فارسی کردن زبانهای برنا مه سازی، گزارش کامپیوتر، شهرپورما ۱۳۶۴، صفحات ۷ - ۵، قاسم جابری پور.
- ۲ - یک راه حل ساده برای مرتب سازی کامپیوتری اطلاعات فارسی - گزارش کامپیوتر اسفند ۶۴، صفحات ۸ - ۳، قاسم جابری پور.
- ۳ - کدا استاندارد برای تبادل اطلاعات در فارسی - سمینار شرکت داده پرداز، آذرماه ۱۳۶۴، دکتر جعفر منشی.

4. Haruyuki Seki, Importance of Standardization Grows with Growth of Demand for Printers, Office Equipment & Products (OEP), May 1985, Volume 14, Serial no. 79, pp. 50-52, DEMPA Publication, INC.

ضمیمه ۱:
مقایسه کننده

سخت افزار

دو منظور

خروجی مقایسه کننده ۴ بیتی شماره ۱ در - شکل ۴ - به نتیجه مقایسه $x_6x_5x_4$ و $y_6y_5y_4$ عیناً به ورودی مقایسه کننده ۴ بیتی بعدی متصل می‌شوند، ولی این خروجی در بعضی موارد می‌بایست اصلاح شود مثلاً اگر $x = 7$ (رمز "چ") و $y = 102$ (رمز "ج") باشد خروجی مقایسه کننده ۱ برابر $EG = 01$ ولی از آنجا که "چ" و

	در y ستون ۵	در y ستون ۷	در y ستون ۶	در y ستون ۴
در x ستون ۵	01	11	10	00
در x ستون ۶				
در x ستون ۷				
در x ستون ۵				

 Q_e/Q_g

$$Q_e = \bar{x}_4\bar{y}_4 + x_4y_4 = x_4 \oplus y_4$$

$$Q_g = x_4y_4$$

شکل ۵: جدول (نقشه‌های کارنا) مقادیر e و g وقتی رمزهای x و y در محدوده الفبا هستند.

"ج" با بستی مساوی تلقی شوند و ورودی مقایسه کننده ۲ در این مورد می‌بایست $eg = 10$ باشد. به عنوان مثالی دیگر 'ض' $x = 1$ و 'پ' $y = 1$ را در نظر می‌گیریم، در این مورد $EG = 00$ یعنی $x < y$ است در موردی که در ترتیب الفبا 'خ' می‌باشد لذا با بستی $eg = 01$ شود. بدیهی است که اگر یکی از x و y یا هر دو در چهار رستون اول - جدول شکل ۲ باشند، $eg = EG$ کفایت می‌کند، به طور کلی برای $x_6y_6 = 0$ (یک یا هر دو x و y در چهار رستون اول) رابطه $eg = EG$ برقرار است و برای $x_6y_6 = 1$ (هر دو x و y در محدوده الفبا یعنی در چهار رستون دوم جدول) eg در تمام موارد برابر EG نخواهد شد. همچنین مقدار C (تعیین کننده معمولی یا دو منظوره بودن مقایسه کننده) در رابطه بین eg ، EG و موثر است به طوری که در $C = 0$ داریم $eg = EG$ و در $C = 1$ eg لزوماً با EG برابر نخواهد بود، به طور خلاصه:

$$D = Cx_6y_6 = 0 \rightarrow eg = EG$$

$$D = Cx_6y_6 = 1 \rightarrow eg = Q_eQ_g$$

که مقادیر Q_e و Q_g با بستی به دست آورده - شوند به طوری که:

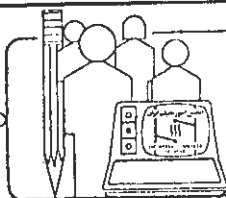
$$e = \bar{D}E + DQ_e$$

$$g = \bar{D}G + DQ_g$$

جدول شکل ۵ مقادیر Q_e و Q_g را در تمام مواردی که x و y در محدوده الفبا (از الف به بعد) قرار دارند مشخص می‌کند و به کمک آن روابط Q_e ، Q_g بر حسب x_4 ، y_4 به دست می‌آیند و از آنجا روابط e و g به شرح زیر نتیجه می‌شوند:

$$e = \bar{C}x_4y_4E + Cx_4y_4(x_4 \oplus y_4)$$

$$g = \bar{C}x_4y_4G + Cx_4y_4(x_4y_4)$$



گزارش کامپیوتر

ماهنامه انجمن انورماتیک ایران

زیر نظر کمیته انتشارات انجمن

نشانی: صندوق پستی ۱۱۹۶ - ۱۴۱۵۵، تهران، ایران

سرپرست کمیته انتشارات و سردبیر: حمیدرضا رهبر

آخرین مهلت قبول مطلب و آگهی: سومین چهارشنبه، ما قبل از انتشار
نقل مطالب با هماهنگی گزارش کامپیوتر با ذکر مأخذ آزاد است

COMPUTER REPORT, Monthly Publication of the Informatics Society of Iran (ISI)
Address: P.O. Box 14155-1196, Tehran, IRAN - Editor: HAMID REZA RAHBAR

کارکرد فارسی در سطوح مختلفی مطرح است و به سبب صرف تشبیه خروجی فارسی نمی توان ادعا کرد که کامپیوتر دارای قابلیت کارکرد فارسی است بلکه تمیل از آن باید انتظار ما را کارکرد فارسی در رابطه با کاربرد مورد نظر معین نمود.

دوم آن که تواما مجموعه ای از حروف لاتین و یک زبان دیگر (مثلا فارسی) در سیستم قرار داشته باشند. در حالت دوم، بسته به نیاز می توان از مجموعه لاتین و یا مجموعه فارسی استفاده کرد. در بسیاری از ریزکا مپیوترهای دوزبانه امروز می توان هر دو زبان را تواما و بطور مخلوط (MIX) به کار گرفت. این قابلیت به اندازه جدول نویسه های مورد استفاده در سیستم (مثلا ASCII) بستگی دارد، اندازه استاندارد این جدول (به ویژه در کامپیوترهای کوچک و کامپیوترهای بزرگ) ۱۲۸ نویسه است که در این صورت در هر لحظه، تنها یکی از مجموعه های فارسی یا لاتین، قابل دستیابی است و نمی توان آنها را به صورت مخلوط به کار برد زیرا حروف زبان دوم نیز از همان کدهای اسکس حروف زبان اول استفاده می کنند. برای مثال، کده ۶ در جدول اول ممکن است معادل A و در جدول دوم معادل "ش" باشد. بنابراین در هر مورد فقط می توان یکی از مجموعه ها را انتخاب نموده و به سیستم معرفی کرد. انتخاب مجموعه (زبان) معمولا با ارسال یک نویسه کنترلی قرار دانی، انجام می گیرد.

اما تقریبا در تمام ریزکا مپیوترها (حتی ریزکا مپیوترهای تک زبانه) اندازه جدول اسکس ۲۵۶ نویسه است که معمولا ۱۲۸ نویسه اول متعلق به زبان لاتین بوده و در ۱۲۸ نویسه دوم علائم و حروف زبان دوم قرار می گیرند. از همین رو است که امکان آن فراهم می شود که بتوان از حروف هر دو زبان تواما و بصورت مخلوط استفاده نمود.

دستگاه های جنوبی آن ردوبدل شود، در درون کامپیوتر در قسمت های ورودی خروجی تحت کد خاصی نگهداری گردد. معمولا کدها می باشد که برای نگهداری این الگوها مورد استفاده قرار می گیرند از نوع ROM که پایداری بوده و محتویات آنها پاک نشدنی است می باشد.

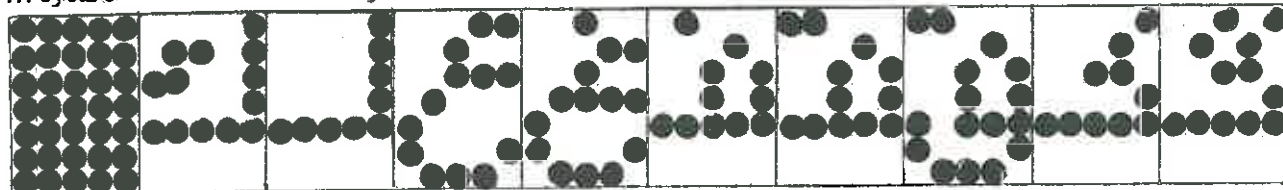
در حقیقت با فشار دادن کلید مربوط به هر علامت بر روی صفحه کلید، کده مربوط به آن علامت وارد کامپیوتر می گردد. سپس هنگام نمایش و یا چاپ، الگوی مربوط به آن بر مبنای این کده از درون حافظه فوق خارج می شود.

کامپیوترها می که دوزبانه کار می کنند در درون این حافظه، الگوی علائم زبان دوم را نیز از قبل جای داده اند. کدهای مربوط به علائم زبان دوم با کدهای علائم زبان اصلی مشابه هستند و یا متما یز می باشند. در حالت اول امکان این که بتوان جملاتی از ترکیب کلمات هر دو زبان داشت یا علاوه بر آن را دویا با سختی مواجه است حال آن که در حالت دوم بین امکان بهر احتیسی قابل ارائه می باشد. به عبارت دیگر، در این رابطه به طور کلی دو حالت عمده در کامپیوترها وجود دارد. اول آن که سیستم صرفا دارای یک مجموعه علائم و حروف (مثلا صرفا لاتین) باشد و

زنچیره دستگا چا پ که علائم زبان لاتین بر روی آنها حک شده بود با گوی یا زنچیره معادل که دارای علائم زبان غیر لاتین بود جایگزین می گردید. برای این منظور قرار دادهای میان برنامه نویسان و متخصصان کامپیوتر و استادان دانشکده های در مورد جدول معادل علائم زبان صورت می پذیرفت تا در هنگام وارد کردن اطلاعات استفاده کنندگان به جای علائم زبان دوم همان علائم معادل در زبان لاتین را وارد نمایند. مثلا حرف "ق" بر روی زنچیره چا پ با حرف "ف" جایگزین شده بود و لذا برای این که "ف" چا پ می شد، استفاده کننده می بایست بر اساس جدول معادلی که در اختیار داشت به جای آن حرف را وارد می نمود.

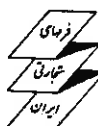
نویسه های فارسی در جدول نویسه ها

اما امروزه مسئله دوزبانه کار کردن اهمیت بیشتری یافته و از این رو کامپیوترها در این مورد مجهز به امکانات بیشتری شده اند. یکی از مهمترین این امکانات وجود شکل حروف و علائم زبان دوم (مثلا فارسی) در درون کامپیوترها است. به طور کلی می بایست از قبل، شکل و الگوی هر علامتی که بخواهد توسط کامپیوتر مورد پردازش قرار گرفته و بین اجزاء گوناگون کامپیوتر و



لطفاً برای مکاتبات به اینجمن فقط از نشانی
تهران
صندوق پستی
۱۴۱۵۵-۱۱۹۶

استفاده کنید



مفتخریم خدمات زیر را برای شما انجام دهیم:

- ۱- تولید انواع فرمهای پیوسته کامپیوتری
- ۲- تولید انواع فرمهای دارای بطریق اسنپ آوت
- ۳- عرضه انواع تجهیزات مغناطیسی کامپیوتری
- ۴- عرضه صندوقهای ذخیره و نشر و انفجار مخصوص نگهداری تجهیزات مغناطیسی کامپیوتری
- ۵- عرضه برجسب های پیوسته کامپیوتری
- ۶- عرضه انواع لوازم اطاق کامپیوتر
- ۷- عرضه دستگاههای تفکیک کننده، جدا کننده و خرد کننده فرمهای پیوسته کامپیوتری
- ۸- عرضه کلاسورهای بایگانی فرمهای پیوسته کامپیوتری
- ۹- عرضه انواع ریبون پرینتر کامپیوتر

فرمهای تجارتنسی ایران، پیشرو در تامین ملزومات کامپیوتری ایران.

خیابان تهران نو، ایستگاه فرودگاه شماره ۳۶۰ تهران ۱۶ صندوق پستی ۴۹۴ - ۱۴۱۱۵

تلفن: ۷۲۱۸۳۱ - ۲۱۳۰۲۹ تلکس: