



فرمت ایرامارک (IRAMARC) برای تبادل اطلاعات فارسی (طرح پیشنهادی)

مجتبی اسدی

در ایران هنوز پایگاهی که بتوان اطلاعات کلیه کتابهای فارسی را به صورت ماشینی در آن ذخیره کرد پدید نیامده است. هر سازمانی به طور پراکنده و جدا از سازمانهای دیگر (از جمله کتابخانه ملی) به ضبط اطلاعات کتابشناختی منابع فارسی و لاتین خود اقدام می‌کند. در این گزارش پیشنهاد شده است که با توجه به گستردگی استفاده از پایگاه MARC و فرم بین‌المللی آن یعنی UNIMARC، این فرمت مبنای کار ایجاد فرمتی ملی برای ضبط اطلاعات کتابهای فارسی قرار گیرد و با توجه به نحوه نامگذاری فرمتهایی که سایر کشورها با استفاده از فرمت MARC برای خود طراحی کرده‌اند نام IRAMARC به آن داده شود.

برای پیگیری مطلب علاوه بر مروری دوباره به ساختار MARC و UNIMARC انواعی از MARC مربوط به کشورهای هندوستان، کره، چین، ژاپن که خط غیرلاتینی دارند مورد بررسی قرار گرفت و علاوه بر آن منابع زیادی دربارهٔ زبانشناسی و خصوصیات خط فارسی و مسائل مربوط به کامپیوتری کردن آن مطالعه شد که از آن میان فقط از منابع معدودی در گزارش حاضر استفاده شده است. جهت آشنایی بیشتر با موضوع مورد بحث یک مورد جستجوی منابع اطلاعاتی به شیوهٔ پیوسته نیز انجام گرفت که منجر به بازیابی منابع زیادی در مورد انواع MARC و زبانشناسی با توجه به خط فارسی و عربی گردید ولی متأسفانه منابع مزبور هنوز در مرحلهٔ سفارش از خارج قرار دارند و امکان استفاده از آنها به جز معدودی که در داخل کشور موجود بودند میسر نگردید. امید می‌رود که با دریافت منابع مورد نظر و انجام جستجوهای آتی، این بررسی از عمق بیشتری برخوردار شده و در نهایت با رفع مشکلات و راهنماییهای صاحب‌نظران بتوان به طراحی کامل فرمت IRAMARC برای کتابهای فارسی و در مراحل بعدی سایر انواع مواد کتابشناختی اقدام کرد.

۱. شناخت مارک (MARC)

۱-۱. تعریف. مارک عبارت است از فهرستتویسی قابل خواندن بوسیلهٔ ماشین و از حروف آغازین عبارت Machine Readable Cataloguing گرفته شده است. مارک در ابتدا برای ضبط و نگهداری اطلاعات کتابشناختی کتابهای موجود در کتابخانه کنگره بوجود آمد و سپس از آن برای تبادل اطلاعات کتابشناختی بین کتابخانه‌های مختلف استفاده شد.



● از نوارهای مارک می توان در فهرست نویسی،

سفارشات کتابخانه ها، رده بندی و کدگذاری

کتابها، و چکیده نویسی استفاده کرد.

۱-۵. یونی مارک. UNIMARC از سرنامهای عبارت Universal Machine Readable Cataloging گرفته شده است و به معنی فهرست نویسی ماشین خوان بین المللی است. این فرمت از مارک زیر نظر اتحادیه بین المللی انجمنهای کتابداری (IFLA) بوجود آمده و توسعه یافته است و هدف از بوجود آوردن آن هموار ساختن راه مبادله اطلاعات کتابشناختی بین مراکز کتابشناسی ملی بوده است. زیرا پس از اینکه کشورهای مختلف با توجه به نیازهای گوناگون خود انواعی از فرمت مارک را بوجود آوردند مسئله تبدیل فرمتهای مختلف به یکدیگر مطرح گردید و کشورهای مختلف ناچار شدند که برنامه های کامپیوتری جداگانه ای برای تبدیل فرمت مورد استفاده خود به هر یک از فرمتهای دیگر مارک بنویسند و این کار چندان مقرون به صرفه نبود. پس از ایجاد فرمت یونی مارک دارندگان فرمتهای مختلف مارک فقط باید برنامه ای برای تبدیل فرمت مارک مورد استفاده خود به فرمت یونی مارک و بالعکس تهیه نمایند و با این کار راه مبادله اطلاعات کتابشناختی بین کتابخانه های ملی که از فرمتهای گوناگون مارک استفاده می کنند هموار گردید.^(۱)

۱-۵-۱. خصوصیات فرمت یونی مارک. فرمت یونی مارک بیشتر بر اساس تبدیل فرمتهای سایر انواع مارک به یکدیگر بنا شده



۱-۲. استانداردهای فرمت رکوردهای مارک. فرمت^۱ رکوردهای موجود بر روی نوارهای مارک براساس قوانین فهرست نویسی انگلوامریکن (AACR)^۲ و استاندارد بین المللی توصیف کتابشناختی (ISBD)^۳ طراحی شده اند و منطبق با استاندارد ملی آمریکا برای مبادله اطلاعات کتابشناختی روی نوارهای مغناطیسی^۴ و همچنین کدهای اسکی (ASCII)^۵ و استانداردهای مؤسسه ملی استاندارد آمریکا (ANSI)^۶ می باشد.^(۱)

۱-۳. موارد استفاده از مارک. از نوارهای مارک می توان در فهرست نویسی، سفارشات کتابخانه ها، رده بندی و کدگذاری کتابها، و چکیده نویسی استفاده کرد. و همچنین در صورت عضویت در سیستم ALERT کتابخانه کنگره از سرویس SDI، شامل اطلاعات رکوردهای تازه منتشر شده کتابها و مجلات و فهرستهای پیش از انتشار CIP می توان استفاده کرد.

۱-۴. انواع مارک. بتدریج که استفاده از مارک گسترش پیدا کرد کشورها و سازمانهای مختلف برای ضبط اطلاعات کتابشناختی منابع خود فرمتهای خاصی از مارک را ابداع کرده و به کار گرفتند. اختلاف فرمتهای مختلف مارک مثل اختلاف قوانین فهرست نویسی با یکدیگر است. همان طور که کتابخانه ها و کشورهای مختلف قوانین فهرست نویسی را برحسب نیازهای خود تغییر داده و به کار می برند، در مورد مارک نیز اگرچه اساس فرمت آن یکی است ولی هر کشور براساس نیازهای خود تغییراتی در فرمت مارک داده است و بر این اساس است که امروزه بیش از ۲۰ نوع فرمت براساس فرمت MARC بوجود آمده است.^۷



است. در این فرمت عناصر توصیفی به ترتیبی که در قوانین ISBD به کار رفته آمده است و امکاناتی هم جهت تغییرات احتمالی آینده در نظر گرفته شده است. در یونی مارک ۹ بخش اصلی (بلوک) در نظر گرفته شده است که اطلاعات مواد کتابشناختی در داخل آنها قرار می‌گیرند و برای هر بلوک ۹۹ بخش فرعی (فیلد) پیش‌بینی شده است. در اینجا بلوکهای اصلی یونی مارک ذکر می‌گردند:

۱-۵-۲. بلوک های اصلی یونی مارک^(۳)

۱ بلوک شناسایی Identification Block

۱ بلوک اطلاعات کدبندی شده Coded Information Block

۲ بلوک اطلاعات توصیفی Descriptive Information Block

۳ بلوک یادداشتها Note Block

۴ بلوک شناسه های ربط دهنده Linkage Entry Block

۵ بلوک عناوین وابسته Related Title Block

۶ بلوک تحلیل موضوعی Subject Analysis Block

۷ بلوک مسئولیت معنوی Intellectual Responsibility Block

۸ بلوک اطلاعات بین المللی International Use Block

۹ بلوک استفاده ملی National Use Block



۲. استفاده از مارک توسط کشورهای دارای خط

غیر لاتینی

فرمت مارک در ابتدا برای نگهداری و مبادله اطلاعات با خط لاتینی طراحی شده است و این ناشی از این امر است که کامپیوترها طوری طراحی شده‌اند که حروف لاتینی را تشخیص بدهند و بدون برنامه‌ریزی جداگانه نمی‌توان از حروف غیر لاتینی در آنها استفاده کرد. به همین دلیل کشورهایی که از خط غیر لاتینی استفاده می‌کنند در تبادل اطلاعات از طریق مارک و یونی مارک دچار مشکلاتی هستند. در بعضی از این کشورها اقداماتی برای برطرف کردن این گونه دشواریها انجام گرفته است که به برخی از آنها اشاره می‌شود:

۱-۲. استفاده از یونی مارک در کتابخانه ملی هندوستان: کشور هندوستان دارای فرهنگها و زبانهای مختلفی است. در این کشور بالغ بر ۱۵ زبان رسمی وجود دارد که عبارتند از: هندی، گجراتی، کانادا، کشمیری، مالایالام، ماراتی، اوریای، پنجابی، سانسکریت، سیندی، تامیل، تلوگو، و اردو. در سال ۱۹۹۰ حدود ۸۰ درصد انتشارات این کشور را موادی که دارای خط غیر لاتینی بوده‌اند تشکیل داده‌اند^(۴). گامهایی که در کتابخانه ملی هندوستان (NLC) برای استفاده از فرمت یونی مارک برداشته شده است عبارتند از:

۱. ایجاد یک جدولی به نام MARCTBLE برای فیلدهای مارک.

۲. ایجاد یک پایگاه اطلاعاتی با استفاده از فرمت مارک به نام MARCRD با کمک برنامه RD از بسته نرم‌افزاری MINISIS.
۳. ایجاد یک پایگاه اطلاعاتی با استفاده از فرمت UNIMARC با کمک برنامه CD از بسته نرم‌افزاری MINISIS.
۴. تبدیل رکوردهایی که در پایگاه MARCRD ذخیره شده‌اند به فرمت UNIMARC با کمک برنامه ISOCONV از بسته نرم‌افزاری MINISIS.
۵. ایجاد یک پایگاه اطلاعاتی برای اسامی مستند مؤلف و تالگان با کمک فایل مستند در MINISIS.

پایگاه MARCRD در حال حاضر تنها شامل رکوردهای تک نگاشت است. سایر انواع مواد از قبیل پیاپیها، مواد سمعی و بصری و مواد کار توگرافی بعداً به آن افزوده خواهند گردید. یکی از دلایلی که فرمت UNIMARC برای کتابخانه ملی هندوستان به عنوان فرمت مبنا انتخاب شده است، قابلیت کار با زبانهای مختلف در این فرمت است و این امر برای هندوستان از اهمیت زیادی برخوردار است (تصویر شماره ۱).

انجمن کتابخانه‌های هندوستان برای حل مشکل خطهای غیر لاتینی آن کشور اقدام به طراحی جدول آوانویسی حروف



ISBN#1	:	d1987 k llengy0103 ba
A001 FIELD	:	1 \$aAmerica's first black woman political writer\$eessays and speeches\$eMaria W. Stewart\$gedited and introduced by Marilyn Richardson
B000 TITLE	:	
A010 LAN	:	0 \$aeng
A020 COCDE	:	\$aUS
B100 IMPRIN	:	\$aBloomington \$eIndiana University Press\$d1987
B150 PAGES	:	\$axxi, 137p.\$cill., map\$d21cm
B250 SERIES	:	1 \$aBlacks in the diaspora
C200 BIBNOT	:	\$aBibliographical references included in "Notes": p.121-132 \$aincludes index
D100 SERNOG	:	1 \$aBlacks in the diaspora
G000 PERISI	:	1 \$aStewart\$bMaria W.
G020 PERSEC	:	1 \$aRichardson\$bMarilyn, ed.
IC00 CALLNO	:	E 326.973 St 49
I010 ACCNUM	:	504604
W900STATUS	:	PACT
W910SCDATE	:	1992-02-20
Z100 ISBN	:	\$a0-253-20446-1\$dJUS\$d22.50
L070 BIBLEV	:	M
F060 TOPIC	:	\$aSlavery\$yU.S.\$xAnti-slavery movements\$xCollections works \$aNegroes in U.S.\$xHistory \$zTo 1863\$xCollections works \$aWomen, Negro\$yU.S.\$xHistory\$z19th century\$xCollections works
W100CATGER	:	AM
F760 DEWEY	:	\$a326.9735v16

تصویر ۱: فرمت MARCRD

موجود در این کشور را از طریق یونی مارک در اختیار سایر کشورها قرار دهد.^(۵)

۲-۲. تأثیر رسم الخط بر فرمتهای مارک کشورهای شرق آسیا^(۶).
از جمله مشکلات مربوط به استفاده از مارک برای کشورهایی که از خط غیر لاتینی استفاده می کنند می توان به مشکلات کشورهای کره، ژاپن، و چین اشاره کرد. کره و ژاپن علاوه بر حروف الفبای مخصوص خود، از الفبای چینی نیز استفاده می کنند که در کشورهای آنها به نام «هنجا» و «کانجی» نامیده می شود. بنابراین بسیاری از اطلاعات کتابشناختی در این کشورها ممکن است با حروف چینی ظاهر شوند. این اطلاعات در هنگام فهرست نویسی ماشینی برای بازیابی کارآمدتر باید به صورت لاتینی شده (Romanized) در آیند.

در ویرایشهای اخیر یونی مارک و USMARC امکان وارد کردن رسم الخطهای غیر لاتینی هم در فیلدهای مربوطه پیش بینی شده است. در کتاب قواعد فهرست نویسی کره ای (Korean Cataloguing Rules) پیشنهاد شده است که سرشناسه های کلیه منابع منتشر شده در زبانهای شرق آسیا به زبان کره ای («هنگول») آورده شوند و منابع منتشره به خط لاتینی به همان خط بیابند، و در ضمیمه قوانین مربوط به KORMARC قواعدی برای آوانویسی کلیه اسامی مربوط به زبانهای شرق آسیا به زبان کره ای آمده است (تصویر شماره ۳).



● فرمت یونی مارک بیشتر بر اساس تبدیل فرمتهای سایر انواع مارک به یکدیگر بنا شده است.

● در ایران تاکنون تلاشی برای ایجاد فرمتی مستقل برای کتابهای فارسی صورت نگرفته است

نموده است. (تصویر شماره ۲) سازمانهای بزرگی همچون شبکه OCLC و RLIN مواد به زبان هندی را به صورت آوانویسی شده به حروف لاتین در بانکهای اطلاعاتی خود ذخیره کرده اند. دفتر کتابخانه کنگره در دهلی نو سیستمی ابداع کرده است که اطلاعات کتابشناختی منابع هندی را به زبان اصلی برمی گرداند. اکنون کتابخانه ملی هندوستان با کمک فرمت MARCRD به شبکه جهانی یونی مارک پیوسته است و می تواند اطلاعات کتابشناختی



अ a	आ ā	ज j = j Telugu	य { y = য় Bengali & Oriya ÿ = য় Bengali, Oriya & Assamese
इ i	ई ī	ञ z	र r
उ u	ऊ ū	झ jh	ڑ { Malayalam & Tamil ڑ
ऋ ṛ	ॠ ṛī	ञ ñ	ल l
ऌ ḷ	ॡ ḷī	ट ṭ	ळ l
ए e	= e N.I.	ठ ṭh	ऌ ḷ
ऐ ai		ड ḍ	ॡ ḷī
औ au	= o N.I.	ण ṇ	ॢ ḷ
क k		त t	थ th
ख kh	क ḱ	द d	ध dh
ग g	ख kh	न n	ञ ñ
घ gh	ग g	ॢ ḷ	ॣ ḷ
ङ ṅ	घ gh	ॣ ḷ	। ḷ
च c = च Telugu	ङ ṅ	। ḷ	॥ ḷ
छ ch	च ch	॥ ḷ	० ḷ

تصویر ۲: جدول آوانویسی حروف

توصیفی (فیلد شماره ۲۰۰ در نظر گرفته شده است. در فیلد مؤلف اثر، فیلدهای شماره ۷۷۰-۷۹۲ به بلوک نقاط دسترسی (Accesspoint) برای اضافه کردن فرم آوانویسی شده نام مؤلف در نظر گرفته شده است. در این قسمت همچنین می توان آوانویسی شده اسم اصلی مؤلف به زبان چینی را هم آورد (تصویر شماره ۵). در گذشته اطلاعات کتابشناختی منابع غیر لاتینی به خط لاتینی (رومی) آوانویسی می شد، ولی این کار باعث بروز مشکلات فراوانی از قبیل تشابه کلمات بخصوص در مورد منابع چینی و ژاپنی و کسره های می گردید. در حال حاضر USMARC و UNIMARC روشی را به کار گرفته اند که براساس آن می توان فیلدهای مختلف را به صورت آوانویسی شده به لاتین و همچنین صورت اصلی آن ثبت کرد. در USMARC فیلد فرعی شماره 6 برای وارد کردن صورت اصلی حروف و فیلد O66 برای معرفی مجموعه های حروف به کار رفته در صورت اصلی اثر در نظر گرفته شده است.

حروف زبانهای چینی و ژاپنی و کراهی (CJK) را در حال



در کشور ژاپن، در چاپ دوم کتاب قواعد فهرست نویسی (Nippon Cataloguing Rules) اشاره شده است که اگر شناسه ای با خطی به جز خط ژاپنی «کانا» و یا خط لاتینی نوشته شده باشد باید آوانویسی به خط «کانا» و یا خط لاتینی (رومی) به آن اضافه شود. JAPAN MARC بر اساس UNIMARC طراحی شده است. در این فرمت فیلدها را به دو قسمت بلوک توصیفی و بلوک نقاط دسترسی (accesspoints) تقسیم کرده اند. اطلاعات در بلوک توصیفی مانند عنوان و مؤلف به همان صورتی که در اصل اثر وجود دارند، وارد می شوند. بنابراین اطلاعات به خط کانجی (خط چینی) بدون هیچ تغییری وارد می گردد، ولی اطلاعات وارد شده در بلوک نقاط دسترسی آوانویسی شده اطلاعات بلوک توصیفی هستند و می توانند به صورتهای «کاناکانا»، لاتینی، و «کانجی» آورده شوند (تصویر شماره ۴).

در تایوان، CHINESE MARC فیلدهایی را برای آوانویسی به صورت لاتینی در نظر گرفته است. در فیلد عنوان اثر، یک فیلد فرعی \$r برای وارد کردن فرم آوانویسی شده در قسمت بلوک

حاضر می‌توان هم به صورت آوانویسی شده به لاتینی و هم به صورت اصلی خود در فیلد 880 مربوط به USMARC وارد نمود (تصویر شماره ۶).

در UNIMARC قسمت خاصی برای ثبت الفبای غیر رومی در نظر گرفته شده است. فیلد فرعی \$6 و \$7 مربوط به الفبا و رسم الخط مواد است. فیلد فرعی \$6 شامل اطلاعاتی می‌شود که با کمک آن می‌توان این فیلد را به سایر فیلدهای رکورد مربوط مرتبط ساخت و فیلد فرعی \$7 شامل کدی برای الفبا و رسم الخط محتویات فیلد اصلی است. این فیلد وقتی که الفبا و رسم الخط با الفبای بکار گرفته شده در فیلد شماره 100 یکی باشد، حذف می‌شود (تصویر شماره ۷).

با توجه به اینکه برای طراحی فرمتی برای وارد کردن اطلاعات منابع فارسی (کتابها) باید به خصوصیات خط فارسی توجه داشت، در فصل بعد خصوصیات خط فارسی مورد بررسی قرار می‌گیرد.

۳. خصوصیات خط فارسی و مشکلات پردازش آن

ابتدا باید به این نکته توجه کرد که خط نشانه محسوس و دیداری نظام ذهنی زبان است و به عنوان ماده‌ای که نظامهای آوایی را در خود منعکس می‌سازد با زبان ارتباط می‌یابد^(۳). به این ترتیب می‌توان گفت که خط باید تحمّل بار معنایی موجود در جلوه‌های مختلف زبان را داشته و بتواند آن را به بهترین نحوی متجلی کند. خط فارسی دارای خصوصیتی است که آن را از خطهای لاتینی و همچنین از خطهای کشورهای شرق آسیا متمایز می‌کند:

۱-۳. شکل حروف فارسی:

۱. حروف ا، د، ذ، ر، ز، ژ، ط، ظ، و معمولاً در همه جا به یک صورت نوشته می‌شوند.
۲. حروف ب، پ، ت، ث، ج، چ، ح، خ، س، ش، ص، ض، ک، گ، ل، م، ن، ی حداقل دو گونه چایی و ماشینی دارند یکی به صورت بزرگ و دیگری به صورت کوچک (که هیچگاه در انتهای کلمه آورده نمی‌شوند).
۳. حروف ف و ق سه گونه دارند: تنها و در آخر (ف، ق)، آغازی (فّ، قّ) و میانی (فء، قء)
۴. حروف ع، غ، ه حداقل به چهار صورت دیده می‌شوند: آغازی: عا، عه، عه، عه، میانی: عا، عه، عه، عه
۵. الصاق انتهای: ع، غ، ه به صورت تنها: ع، غ، ه حروف متصل و منفصل: حروف متصل آنهایی هستند که از طرف چپ به حروف دیگر نمی‌چسبند ولی از طرف

100	\$0\$aSurname\$bForename (in Hangeul)
245	X1\$aTitle\$cName of the Author (in Hanja)
600	Personal Name Subject Heading (in Hangeul)
740	Added Entry -- Title Traced differently (in Hangeul)

تصویر ۳: نمونه‌ای از KORMARC

251	\$ATitle\$FAuthor (in Kanji)
551	\$ATitle(in Gatakana)\$XTitle(in roman)\$BTitle(in Kanji)
650	\$APersonal Name Subject Heading(in Gatakana)\$XPersonal Name Subject Heading(in roman) \$BPersonal Name Subject Heading(in Kanji)
751	\$AAuthor(in Gatakana)\$XAuthor(in roman)\$FAuthor(in Kanji)

تصویر ۴: نمونه‌ای از JAPANMARC

200	XX\$aTitle(Chinese)\$/Author(Chinese)\$/Title (romanization)
700	XX\$aAuthor-Surname\$bAuthor-Forename (in Chinese)
770	XX\$aAuthor-Surname\$bAuthor-Forename (romanization)

تصویر ۵: نمونه‌ای از CHINESEMARC

066	= c\$]
245	00 =6580-02#aTitle[romanized]
250	=6880-22#aEdition[romanized]
650	0 =aSubject(no alternative representation]
850	00 =6245-02/#aTitle[CJK data]
880	=6250-22/#aEdition[CJK data]

تصویر ۶: نمونه‌ای از USMARC

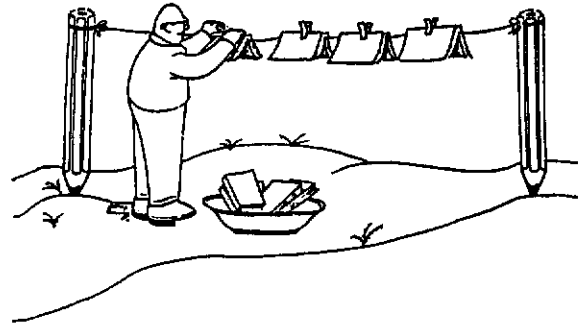
200	1#6a01\$a[Title in Korean characters]
200	/1#6a01\$b#a\$Title[romanized]

تصویر ۷: نمونه‌ای از UNIMARC



الف: نظامی که در آن اسامی و اصطلاحات زبانهای عربی، فارسی، ترکی، عثمانی، اردو (یعنی زبانهایی که با الفبای عربی نوشته می‌شوند) به صورتی واحد می‌آید. از این نظام در دایرةالمعارف اسلام و دانشنامه ایران و اسلام استفاده شده است. حرف‌نگاری در این نظام در صفحه ۴۰ دانشنامه ایران و اسلام به این صورت آورده شده است^(۱۰):

حرف‌نگاری ابرامارک (IRAMARC)



راست می‌توانند به حروف معینی متصل شوند. این خصوصیت یکی از مشکلات رسم‌الخط فارسی است. این گروه عبارتند از:

ا، د، ذ، ز، ژ، و بقیه حروف از نوع منفصل‌اند.

۶. بسیاری از حروف فارسی پایه مشترک دارند و اختلاف آنها با گذاردن نقطه‌هایی در بالا یا در زیر حرف مشخص می‌شود. مثلاً: ب، پ، ت، ث / ج، چ، ح، خ، د، ذ، ر، ز، ژ / س، ش / ص، ض / ع، غ / ف، ق

۷. علائم حروف فارسی عبارتند از: فتحه، کسره، ضمه، تشدید، سکون، مد و تونینهای سه گانه، ی، ِ^(۸)

صامتها

ر	r	غ	gh
ز	z	ف	f
س	s	ق	k
ش	sh	ک	k
ص	s	ل	l
ض	d	م	m
ط	t	ن	n
ظ	z	ه	h
ع	ʿ	و	w
		ی	y

مصوت‌های بلند

آ	ā
و	ū
ی	i

مصوت‌های مرکب

او	aw
ای	ay
یی	yy
وو	uw

مصوت‌های کوتاه

ا	a
و	u
ی	i

نقل حروف مخصوص فارسی و ترکی وارد به لاتین

پ	p	ژ	zh	ب‌ط	ʿ
چ	ç	ک یا ل	h	د‌ط	aʿ

مثال: چند کلمه که با این روش حرف‌نگاری شده‌اند:

qawm	قوم	Husayn	حسین
Abū'l-Khayr	ابوالخیر	kitab	کتاب
tayār	طیار	gharīb	غریب
hūrī	حوری	qarar	ضرر
thulth	ثلث	dhawq	ذوق
Zālim	ظالم	shūsh	شوش
djabr	جبر		
samt	سمت		



۲-۳. آوانویسی (حرف‌نگاری) حروف فارسی

نوشتن کلمات و جملات یک زبان با الفبای زبان دیگر را حرف‌نویسی یا حرف‌نگاری (Transliteration) می‌گویند. معمولاً در حرف‌نگاری نوشتن کلمات و جملات مکتوب مورد نظر است. ممکن است بخواهیم بدون توجه به تلفظ تنها «صورت مکتوب» زبانی را با رسم‌الخط زبان دیگر نشان بدهیم و در صورت لزوم بخواهیم که رشته مورد نظر خود را به صورت اول آن یعنی به زبان اصلی برگردانیم و آن را عیناً بازسازی کنیم. در طریقه سوم، که التقاطی (eclectic) نام دارد در عین حال که می‌خواهیم حرف‌نگاری به طریقی باشد که بازسازی نوشتار زبان اصلی میسر گردد، در حد امکان به نشان دادن تلفظ کلمات نیز توجه می‌شود. این شیوه‌ای است که در حدفاصل صوت‌نگاری و حرف‌نویسی مطلق قرار می‌گیرد^(۹).

دو نظام حرف‌نگاری توسط مستشرقین در ضبط کلمات فارسی به کار گرفته شده است:

ب: نظام دوم برای حرف نویسی کلمات فارسی توسط ناصر شریفی تنظیم شده است:

sh	ش	-	ا
s	ص	b	ب
z	ض	p	پ
t	ط	t	ت
z	ظ	ṣ	ث
'	ع	j	ج
gh	غ	ch	چ
f	ف	h	ح
q	ق	kh	خ
k	ک	d	د
g	گ	z	ذ
l	ل	r	ر
m	م	z	ز
n	ن	zh	ژ
v	و	s	س
h	ه		
y	ی		

مصوتها و مصوت‌های مرکب:

ā	آ	ای	اَ
ī	آ	یی	اِ
ū	و	وِ	اُ
ey	ئی	ئی	اِی

نشانه‌ها

,	،
ˆ	ˆ

کلمات قبلی با این نظام به این صورت نوشته می‌شوند:

qawm	قوم	Husayn	حسین
Abū al-Keyr	ابوالخیر	kitāb	کتاب
ṭayyār	طیار	gharīb	غریب
ḥūrī	حوری	zarar	ضرر
ṣofā	ثلث	Zowb	ذوب
Zālem	ظالم	shūsh	شوش
jabr	جبر		
samt	سمت		

مرکز اسرار کلامی (IRAMARC)

در رابطه با این دو شیوه نکاتی به نظر می‌رسند که عبارتند از:

۱. در هر دو شیوه اصل مکتوب زبان عربی و فارسی را به آسانی می‌توان بازسازی کرد.
 ۲. هر دو شیوه تلفظ را تا حدی نشان می‌دهند.
 ۳. شیوه اول در اصل برای زبان عربی تهیه شده ولی ضبط کلمات فارسی هم با آن انجام گرفته است، ولی شیوه دوم مخصوص حرف‌نویسی زبان فارسی است.
- در زبان فارسی جدید هنوز شیوه‌ای منظم و علمی برای حرف‌نویسی کلمات وضع نشده است و در نگارش اسامی خارجی به رسم الخط فارسی هر کس تا حد زیادی سلیقه خاص خود را به کار می‌برد.

۳-۳. مشکلات ماشینی کردن خط فارسی

۱. همانطور که اشاره شد کامپیوترها را برای خواندن خطوط لاتینی طراحی کرده‌اند و خط فارسی برخلاف خط لاتین از راست به چپ نوشته می‌شوند. در برنامه‌ها به علت عدم وجود سخت‌افزاری که بتواند از راست به چپ عمل کند احتیاج به نوشتن جداگانه برنامه‌های جابجاکننده وجود دارد.
۲. همچنین کلمات لاتین با فاصله‌ای که بعد از آخرین حرف هر کلمه گذاشته می‌شود براحتی از یکدیگر تفکیک می‌شوند ولی در فارسی حروف کلمات به دو دسته متصل و منفصل تقسیم می‌شوند و گاهی حروف به طور متصل و گاهی به طور منفصل کلمات را به وجود می‌آورند، بنابراین در اکثر موارد بودن یک فاصله خالی یا عدم آن به هیچوجه نشان‌دهنده پایان یک کلمه نیست.

۳. چون بعضی از حروف فارسی بیش از یک صورت نوشته می‌شوند ناچاریم برای نگهداری کلیه حالت‌های حروف بیش از ۳۲ کاراکتر را در کامپیوتر به این کار اختصاص دهیم.

۴. در زبان لاتین می‌توان تمام حروف را بین دو خط موازی محاط کرد ولی در زبان فارسی حروف تشکیل‌دهنده یک کلمه در سطوح مختلف نوشته می‌شوند و برای محاط کردن آن به حداقل سه خط موازی نیازمندیم: مثال ف ق آ

۵. حروف لاتین دارای عرض مساوی هستند در صورتیکه حروف فارسی از این نظر ناهماهنگی زیادی دارند مثلاً حروف اه، ق، ک دارای عرض‌های مختلفی هستند و باعث بروز مشکلاتی می‌شوند^(۱۲).





۴. پیشنهاد فرمت ایرامارک (IRAMARC) برای کتابهای فارسی

با توجه به مطالبی که در فصول پیشین عنوان شد ملاحظه می‌گردد که بیش از بیست نوع فرمت MARC در حال حاضر در کشورهای مختلف براساس نیازهای خاص آنها طراحی شده و مورد استفاده قرار گرفته است.

در ایران تاکنون تلاشی برای ایجاد فرمتی مستقل برای کتابهای فارسی صورت نگرفته است و به نظر می‌رسد که انجام آن فرصتی بوجود خواهد آورد تا بتوانیم کلیه اطلاعات کتابشناختی کتابهای فارسی را به صورت ماشینی خواه درآورده و در اختیار سایر کشورهای جهان قرار دهیم.

برای این کار لازم است که پس از طراحی فرمت مربوط به ایران که آن را ایرامارک می‌توان نامید، برنامه‌هایی نوشته و طراحی شوند که اطلاعات کتابشناختی کتابهای فارسی موجود در پایگاه اطلاعاتی ایرامارک را به فرمت یونیمارک تبدیل کرده و در اختیار کشورهای دیگر قرار دهد.

در این زمینه باید به خصوصیات خط فارسی هنگام ضبط در کامپیوتر توجه شود و همچنین فرمت کشورهای کره، چین و ژاپن از این نظر که خط آنها نیز مثل خط فارسی غیر لاتینی است مورد توجه دقیق قرار گیرند.

سخت‌افزار مورد نیاز پس از تعیین حجم اطلاعات کتابشناختی کتابهای فارسی تعیین خواهد گردید.

برای تعیین خصوصیات کتابشناختی مورد نیاز برای فرمت ایرامارک می‌توان از کتابشناسی ملی ایران به عنوان مبنای استفاده کرد.

منابع و مراجع:

۱. مجتبی اسدی انجیله، معرفی سیستم‌های MARC و UNIMARC و بررسی کاربرد UNIMARC (پایان‌نامه فوق‌لیسانس کتابداری، دانشکده علوم تربیتی دانشگاه تهران، ۱۳۶۹)، به راهنمایی آقای دکتر عباس حرزی، ص ۵.

۲. همان، ص ۳۳
۳. همان، ص ۳۵

4. UmaMajumder, "Implementation of UNIMARC at India's National Library" in *International Cataloguing & Bibliographic Control* [21(2)1992] p.22
5. Ibid. p.23
6. Jeong Pil-Mo and OhDong Geun, "The Influence of Pronunciation and Scripts On the East-Asian MARC Formats: in *International Cataloguing & Bibliographic Control* (20,4,1991)p.56-60
۷. مهری باقری، مقدمات زبانشناسی رشته زبان و ادبیات فارسی (تهران: دانشگاه پیام نور، ۱۳۷۰)، ص ۱۱۳
۸. لطف‌الله یارمحمدی، درآمدی به آشناسی (تهران: مرکز نشر دانشگاهی، ۱۳۶۴)، ص ۱۸-۱۶
۹. همان، ص ۳۵
۱۰. همان، ص ۳۶
۱۱. همان، ص ۴۰
۱۲. ناصر آقاموسی ترکمانی، نمایش حروف و کلمات فارسی (پایان‌نامه لیسانس مهندسی علوم، دانشکده ریاضی و علوم کامپیوتر دانشگاه صنعتی [شریف]، ۱۳۵۵)، ص ۲.

یادداشتها:

۱. شکل‌بندی اطلاعات در کامپیوتر

2. Anglo American Cataloging Rules
3. International Standard for Bibliographical Description
4. American National Standard Format for Bibliographic Information Interchange on Magnetic Tape
5. American Standard Codes for Information Interchange.
6. American National Standard Institute

۷. نام این فرمتها به قرار زیر است:

MARC, IBERMARC, CANMARC, SINGMARC, AUSMARC, USMARC, LCMARC, OCLCMARC, UKMARC, UNIMARC, BRIMARC, TELEMARC, REMARC, CATMARC, DMARC, AMRS, ANNAMARC, COMARC, DOBISMARC, MELITMARC, NORMARC, JAPANMARC, KORMARC, MARCRD, CHINESE MARC.