ڈاکٹر حافظ صفوان محمد چوہان

کمپیوٹر کے لیے اُردو حروف

Urdu Letter for Computer

Abstract:

Dr. Hafiz Safwan Muhammad Chohan, Consultant, Urdu Informatics, National Language Authority, Islamabad.

The conventional Urdu script is Perso-Arabic in style & structure. Diacritic marks were added to the ghost characters initially for the ease of foreign speakers and users of the language. This article deals with the fact that characters and diacritic marks are two different things; and if they are treated as such, machine readable Urdu can benefit a lot thanks to this thesis.

In the recent past, the Unicode Consortium has allotted alphabet status to the diacritic marks on the proposal of National Language Authority of Pakistan where R&D on Urdu language is underway at the Centre of Excellence for Urdu Informatics. Add to many other areas of informatics like provision of writing almost all scripts of Pakistani languages in complete machine readable format, this approach is going to advance a sharp twist in the development of Urdu OCR.

تعارف

زبان اور بولی آوازوں کا مجموعہ ہوتی ہے۔ حروف شناختی علامات ہیں اور اِن کی حیثیت اُن لباسوں کی ہے جو اِن آوازوں کی شناخت کے لیے مقرر کیے گئے ہوتے ہیں۔ اِسی لیے ایک ہی آواز کے لیے مختلف زبانوں میں مختلف شناختی علامات ہوتی ہیں۔ جس زبان یا بولی کے حروفِ تہجی نہیں ہوتے اُسے ideograph کہتے ہیں۔

ہر زندہ زبان کا رسم الخط بھی زندہ ہوتا ہے، یعنی اُس میں اِصلاحات ہوتی رہتی ہیں۔ نئی آوازوں کے لیے نئے حروفِ تہجی بھی بنتے ہیں؛ یا کبھی کسی حرفِ تہجی کی شکل میں جزوی

تبدیلی بھی کر لی جایا کرتی ہے۔ رسم الخط میں اصلاح (یا کبھی کبھی کسی تبدیلی) کی ضرورت زیادہ تر دخیل الفاظ کی وجہ سے پڑتی ہے۔ یا پھر جب کسی زبان کے استعمال کرنے والوں کا دائرہ بہت پھیل جائے اور ایسے لوگ کسی زبان کو استعمال کرنے لگیں جن کی وہ مادری زبان نہیں ہے تب بھی بڑے پیمانے پر اِصلاحات کرنا ضروری ہو جاتی ہیں۔ ہمارے تناظر میں عربی زبان اِس دعوے کی سب سے بڑی، زندہ اور عام مثال ہے۔

تحریر میں نقطوں اور شوشوں کی ابتدا: تحریک، تاریخ اور جائزہ

عربی رسم الخط میں بنیادی چیز مختصر نویسی ہے۔ ''مجرد حروف کے دائروں وغیرہ کو چھوڑ کر صرف جزوِ لازم سے کام لے کر حروف کو باہم ملانے''[۱] کا طریقہ دنیا کے کسی بھی اور رسم الخط میں موجود نہیں ہے۔ جب غیر عرب لوگ عربی پڑھنے لگے یعنی جب قرآنِ پاک دنیا بھر میں پھیلنا شروع ہوا تو عربی رسم الخط میں بڑے پیمانے پر ضروری اِصلاحات کی گئیں تا کہ پڑھنے والوں کو سہولت رہے۔ یہ اِصلاحات کچھ تو خود عرب خلفائے اسلام نے کیں اور کچھ علاقائی ضروریات کے تحت اِن آخذین یعنی recipients نے موقع بموقع کیں۔ پہلی اِصلاحات کی مثال نقطوں کا عربی کے حروف پر لگایا جانا ہے جب کہ دوسری قسم کی اِصلاحات کی مثال مختلف قسم کے شوشے اور کش وغیرہ ہیں جو ہماری پاک و ہند کی زبانوں میں بہت کثرت سے پائے جاتے ہیں۔

اردو کے روایتی رسم الخط (جسے اِس مقالے میں آئندہ ''اردو رسم الخط'' لکھا جائے گا) کا خطِ کوفی سے گہرا تعلق ہے۔ ''خطِ کوفی دراصل خطِ نبطی، خطِ سطرنجیلی، خطِ حمیری اور خطِ حیرہ کی آمیزش اور سنبھلی ہوئی صورت تھی۔ دوسرے مقامات کے علاوہ یہ خط حجاز اور حیرہ (کوفہ کا پرانا نام) میں بھی لکھا جاتا تھا۔ حرب ابنِ امیہ اِسے حیرہ سے سیکھ کر آئے اور جزیرۃ العرب میں پھیلایا۔''[۲] حضورِ پاک صلی اللہ علیہ وسلم کی بعثت کے وقت حجازِ مقدس میں عام طور سے خطِ کوفی ہی رائج تھا۔ آپ صلی اللہ علیہ وسلم نے مختلف بادشاہوں اور سرداروں کو جو خطوط لکھوائے وہ خطِ حمیری اور خطِ حیرہ کی اِسی محتاط روش یعنی خطِ کوفی میں لکھے گئے تھے۔[۳] ابتداءً اِس خط میں اِعراب اور نقطوں کا چلن نہیں تھا۔ ابوالاسود دوئلی نے تقریباً سنہ ۵۰ ھ میں نقطے ایجاد کیے۔ لیکن یہ نقطے حروف کی شناخت کے بجائے اِعراب کے طور پر استعمال ہوتے تھے۔[۴]

جب اسلام مصر اور ایران تک پھیل گیا تو اکثر حروف کے ہم صورت ہونے کی وجہ سے قرآنِ مجید کی قرات میں مشکل ہونے لگی۔ مراسلت میں بھی کئی طرح کے مغالطے پیدا ہونے لگے۔ چنانچہ خلیفہ عبدالملک بن مروان نے عراق کے گورنر حجاج بن یوسف کو رسمِ خط کی اصلاح کا حکم دیا، جس نے نصر بن عاصم کو یہ کام تفویض کیا۔ ۵

نصر بن عاصم نے اِعراب کے لیے نقطوں کے استعمال کو بدستور قائم رکھا لیکن اب اِن نقطوں کے لیے صرف قرمزی رنگ استعمال کرنے کی تجویز پیش کی۔ ہم شکل حروف کی تخصیص اور تشخیص کے لیے بھی اُس نے نقطے ہی ایجاد کیے لیکن اِن نقطوں کے لیے سیاہ رنگ لازم قرار دیا۔ خطِ کوفی میں یہ ترتیب تقریباً چالیس سال تک قائم رہی حتیٰ کہ عبدالرحمٰن خلیل بن احمد عروضی نے اِعراب کی شکلیں وضع کر کے اُنھیں سیاہ نقطوں سے الگ کر دیا۔ ۶

خطِ کوفی اپنی سادہ روش اور واضح ہونے کی وجہ سے لوگوں میں مقبول ہوا اور تقریباً چھٹی صدی ہجری تک مستعمل رہا۔ اِس دوران میں اِس میں بعض تبدیلیاں بھی ہوئیں۔ دوسری صدی ہجری میں اِس میں زاویے نمایاں ہوگئے۔ اِس کے بعد اِس خط نے قوسی شکل اختیار کرلی اور رفتہ رفتہ تزئینی اور آرائشی ہوتا چلا گیا اور بالآخر رواج سے خارج ہوگیا۔ پھر اِس کی جگہ نسخی خط نے لے لی۔ ۷ اِس تبدیلی کے باوجود اِعراب، نقطے اور کش البتہ اپنی جگہ پر قائم رہے۔ اِن کی شکل یہ تھی کہ نقطے تو مستقلاً لگائے جانے لگے جب کہ اِعراب ضروری مقامات پر اور کہیں کہیں صرف آرائش کی خاطر استعمال کیے جاتے تھے۔ (مدتوں بعد ''خطِ کوفی جدید'' کے نام سے خطِ کوفی کا احیا ہوا لیکن اِس کے چپٹے حروف کو خطِ کوفی کی روش سے کوئی تعلق نہیں۔)

نسخ کے آرائشی ہوتے چلے جانے کے ساتھ ایران میں، جہاں نسخ اپنے پاؤں جما چکا تھا، دفتری کاموں کے لیے ''تعلیق'' خط اُبھرنا شروع ہوا۔ اِس کی کششوں میں کسی قدر ترچھاپن اور دائروں میں گراوٗ نمایاں تھا اور ایک کش کے سرے دوسرے کے حرفوں سے مل سکتے تھے، یعنی کئی حرف ایک دوسرے سے پیوستہ کر کے لکھے جاسکتے تھے۔ یہ خط زود نویسی کے کام آتا تھا۔ چنانچہ یہ خط بہت مقبول ہوا اور فرامین اور سرکاری خط کتابت کے لیے چلن دار ہوگیا۔ ۸ خواجہ ابوالعالی بک نے خطِ تعلیق میں انقلابی اِصلاحات کیں۔ فارسی کی مخصوص آوازوں (پ، چ، ژ اور گ) کے لیے حروف ایجاد کیے اور اِس مقصد کے لیے تین تین

نقطے وضع کیے۔ ابتدا میں گ پر بھی دو کشوں کے بجائے تین نقطے لگائے جاتے تھے لیکن لسانی اور جمالیاتی ذوق سے ہم آہنگ نہ ہونے کے باعث یہ روش رفتہ رفتہ ترک کر دی گئی اور گ پر دو کش لگائے جانے لگے۔9

نسخ میں حروف کا پھیلاؤ افقی ہوتا ہے جس سے لکھنے میں روانی پیدا ہوتی ہے۔ اِس خط نے سادگی، حسن اور حروف کی واضح اشکال کے باعث ایک مستقل حیثیت اختیار کر لی تھی اور عربی کے ساتھ ساتھ کئی دوسری زبانوں مثلاً فارسی، سندھی، پشتو اور اردو کے لیے استعمال ہونے لگا تھا۔ ساتھ ہی ساتھ تعلیق کی مقبولیت کا دائرہ وسیع ہوا اور اِن دونوں خطوں کے اتصال سے ''نستعلیق'' بنا۔ اِسے اپنے زمانے کے دوسرے خطوں کی ترقی یافتہ شکل کہا جاسکتا ہے۔ لکھائی میں رواں تر اور حروف کے دائروں کے گول ہونے کے فطری رجحان کی وجہ سے یہ خط ایران کی حدود سے نکل کر اَکناف میں پھیل گیا۔10

لیکن عام تحریر کی رفتار نستعلیق کی نشست میں داخل اہتمام کی متحمل نہیں ہوئی۔ بدیں وجوہ بارہویں صدی ہجری کے اوائل میں مرتضیٰ خان شاملو نے نستعلیق سے ایک اور خط اخذ کیا جسے خطِ شکستہ کا نام دیا گیا۔ اِس سے دفتری خط کتابت اور سرکاری دستاویزات میں بہت سہولت ہوگئی۔ خطِ شکستہ میں حروف اور اُن کے دائرے اَدھورے ہوتے ہیں جنھیں لکھتے ہوئے وقت کی خاصی بچت ہوتی ہے، یعنی یہ وقت کو اَدھیا دیتا ہے۔ زود نویسی کے لیے انتہائی کارآمد ہونے اور کئی دوسری اِفادی ضرورتوں اور جمالیاتی جاذبیت کے علی الرغم آج کی عام تحریر میں نقطوں اور شوشوں کی حد تک جو بد نظمی اور انتشار پایا جاتا ہے وہ خطِ شکستہ ہی کے سبب سے ہے۔ یہ خط اگرچہ رواج سے اُٹھ چکا ہے لیکن اگر اِسے متعین اصولوں کے تحت لکھا جائے تو تیز اور گھسیٹ خطِ شکستہ بآسانی خطِ نستعلیق کی طرح پڑھا جاسکتا ہے۔11 شوشوں اور نقطوں کی درست نشست میں باقاعدہ اہتمام کا رجحان مفقود ہونے کی وجہ یہی خط ہے۔

اردو رسم الخط: ابتدا، حروف، نقطے، شوشے

اردو کے آغاز کے ساتھ ہی برِعظیم پاک و ہند میں مسلمان اردو کو فارسی (نسخ و نستعلیق) جب کہ ہندو دیوناگری رسم الخط میں لکھنے لگے۔ اِس کی وجہ، زیادہ تر، یہی تھی کہ ناگری خط میں

عربی اور فارسی کی بعض مخصوص آوازیں (مثلاً خ، ذ، ز،ض، ظ، اور ع وغیرہ) ادا کرنے کے لیے حروف موجود نہیں تھے۔ چنانچہ عربی اور فارسی کے بے شمار الفاظ ایسے تھے جو مسلمانوں اور ہندوؤں کی بولی جانے والی اردو میں مستعمل تھے لیکن اُنھیں ناگری میں لکھنا ممکن نہ تھا۔ دوسری جانب فارسی میں ہندی کی مخصوص آوازیں (مثلاً ٹ، ڈ، ڑ، بھ، پھ، تھ، ٹھ، جھ، چھ، دھ، ڈھ، ڑھ، کھ، گھ، لھ، نھ وغیرہ) ادا کرنے کے لیے حروف موجود نہیں تھے۔ اردو چونکہ اِن تینوں زبانوں کے مرکب سے پیدا ہوئی تھی لہٰذا ضروری ہوا کہ اِس میں اِن تینوں زبانوں کی آوازیں ادا کرنے کے لیے علامتیں موجود ہوں۔ چنانچہ فارسی حروف میں شکلوں کی قبیلہ وار تقسیم سے فائدہ اُٹھا کر دو نئے نشانات وضع کیے گئے جن کے ذریعے ہندی کی مخصوص آوازیں ادا کرنے کے لیے نئے حروف بنانا ممکن ہوگیا۔ چھوٹی طوئے ''ط'' کی علامت کو ب، د اور ر کی قبیل کے الفاظ پر لگا کر ٹ، ڈ، اور ڑ وضع کیے گئے۔ ''ھ'' کی علامت سے بھ، پھ، تھ، ٹھ، جھ، چھ، دھ، ڈھ، ڑھ، کھ، اور گھ بنا لیے گئے۔ اردو حروف کے لیے یہ اضافہ اِتنا بے ساختہ اور فطری تھا کہ اجنبیت اور مصنوعی پن کا احساس نہیں ہوتا۔ اردو کے لیے اِس نئی ابجد کو اختیار کرلیا گیا اور اِسے خطِ نستعلیق میں لکھا جانے لگا۔ حروف کے اِضافے کی وجہ سے ہم اِسے فارسی یا عربی نسخ کہہ کر انصاف نہیں کرتے۔

یہاں اِس بات کا ذکر ضروری ہے کہ ''ھ'' کی علامت بھی اصلاً عربی ہی سے درآمدہ ہے۔ ہائے ہوز— جو اردو، فارسی اور ہندی میں آ کر کبھی ہائے کہنی دار میں بدل جاتی ہے اور جس کی اصل شکل ہ ہے— عربی کے بعضے قدیم رسوم الخط میں ھ بھی لکھی جاتی رہی ہے، اور آج تک لکھی جاتی ہے۔

اردو کا قدیم اور جدید رسم الخط— مشینی دور میں

کاغذ کے دور سے سائبر دور میں داخل ہوجانے کے بعد دنیا بھر کی زبانوں نے اپنے رسوم الخط میں ایسی تبدیلیاں کی ہیں جن کی بنا پر رسم الخط کو کمپیوٹر اور اِنٹرنیٹ پر پڑھا جا سکے۔ یہ بھی ضروری ہے کہ یہ پڑھا جانا صرف آنکھوں کے دیکھنے کی حد تک نہ ہو بلکہ کمپیوٹر اور اِنٹرنیٹ اِس لکھے کو ''سمجھ'' بھی سکیں۔ اِس چیز کو آسان الفاظ میں رسم الخط کا ''مشین ریڈ ایبل'' ہونا

کہتے ہیں۔ اردو والوں نے بھی، گو کچھ دیر کر کے، اِس طرف توجہ کی ہے اور اپنے رسم الخط میں ایسی بنیادی تبدیلیاں کی ہیں جن کی وجہ سے مشین ریڈایبل اردو کی طرف سفر شروع ہوگیا ہے۔ اردو کے بنیادی حروفِ تہجی کی تعداد ۳۸ ہے جو مولوی محمد اسمٰعیل میرٹھی کے قاعدے میں موجود ہے:

ا ب پ ت ٹ ث ج چ ح خ د ڈ ذ

ر ڑ ز ژ س ش ص ض ط ظ ع غ ف

ق ک گ ل م ن و ہ ھ ء ی ے

دورِ جدید کی اردو اور اِس اردو کو مشین ریڈایبل بنانے کے لیے مقتدرہ قومی زبان اسلام آباد نے جو معیارات طے کیے ہیں، اُن کی روشنی میں حروفِ تہجی کی تعداد اب ۵۸ ہوگئی ہے۔ یہ فہرست ملاحظہ کیجیے: http://www.nla.gov.pk/beta/images/alphabetsfull.gif

مشینی دور اور رسوم الخط کی معیار بندی: یونی کوڈ کنسورشیم

مشینی دور میں داخل ہوجانے کے بعد اِس بات کی ضرورت محسوس ہوئی کہ ہر زبان کے لیے ایک کوڈ پلیٹ مہیا ہو اور لکھنے والا اِسی کے مطابق لکھے۔ اِس معیار بندی کے فائدے ظاہر و باہر ہیں۔ سب سے بڑی بات یہ ہے کہ ایک کمپیوٹر سسٹم پر لکھی گئی زبان کسی دوسرے کمپیوٹر سسٹم پر بھی یکساں سہولت سے اور کچھ کھوئے بغیر پڑھی جاسکتی ہے۔ (یہاں زبان سے مراد کمپیوٹر میں داخل کیا گیا مواد ہے۔) دوسرا بہت بڑا فائدہ یہ ہے کہ ایک کوڈ پلیٹ کی بنیاد پر لکھے گئے مواد کی کسی بھی دوسری کوڈ پلیٹ کے لیے انتقالِ حرفی (Transliterate) آسان اور ممکن ہے۔ وغیرہ وغیرہ۔ اِس بنیادی ضرورت کو یونی کوڈ (www.unicode.org) نے پورا کیا۔ اِس کا مختصر تعارف یہ ہے کہ یہ ہر حرفِ تہجی کو ایک منفرد نمبر (unique number) تفویض کرتا ہے جو کسی بھی پلیٹ فارم، کسی بھی پروگرام اور کسی بھی زبان میں کام کرنے کے لیے یکساں رہتا ہے۔ اِس کی ترتیب یہ ہے کہ کمپیوٹر بنیادی طور پر صرف نمبرات سے واسطہ رکھتا ہے۔ جو بھی ہندسہ یا حرف کمپیوٹر پر لکھا جاتا ہے، وہ اِسے کوئی نہ کوئی نمبر دے کر اپنے پاس محفوظ کرلیتا ہے اور پھر اُس پر عمل (processing) کرتا ہے۔ یونی کوڈ سے پہلے کوڈنگ کے لیے سیکڑوں طرح کے نظامات تھے۔ لیکن کوئی بھی ایک نظام سارے الفاظ کی سہار نہیں رکھتا

تھا۔ مثلاً صرف یورپین یونین ہی کو اپنی تمام زبانیں کمپیوٹر پر پڑھے جانے کے قابل بنانے کے لیے کوڈنگ کے کئی ایک نظامات درکار تھے۔ اِسی پر بس نہیں بلکہ کسی بھی ایک زبان۔۔ جیسے مثلاً انگریزی۔۔ کے لیے کوڈنگ کا کوئی ایسا نظام موجود نہیں تھا جو تمام حروف، رموزِ اوقاف اور تکنیکی اِصطلاحات وغیرہ کو مناسب انداز میں برت سکتا۔ یونی کوڈ نے اِس بے ہنگم پن کو دور کر کے معیار بندی کو فروغ دیا ہے۔۱۲

مشینی دور اور اردو رسم الخط کی دشواریاں: معیار بندی کی طرف سفر

اردو رسم الخط کی بنیادی دشواریوں کو ہم دو حصوں میں تقسیم کر سکتے ہیں۔

۱۔ اتصالی مشکلات: وہ دشواریاں جو اَشکالِ حروف (یعنی تناظر کے بدل جانے سے ایک ہی حرف کی کئی کئی شکلوں کا وجود میں آجانا)، اتصالِ حروف، شوشوں، مرکز اور اِعراب وغیرہ سے متعلق ہیں۔

۲۔ صوتی مشکلات: وہ دشواریاں جو مشتبہ الصوت حروف اور الفاظ وغیرہ سے متعلق ہیں۔۱۳

اردو میں ایک ہی لفظ کی کئی کئی شکلیں بننا بہت عام ہے۔ مثلاً ب ہی کی کم سے کم ۱۴ شکلیں موجود ہیں۔ چنانچہ ایک کوڈ پلیٹ کی بنیادی ضرورت سے اردو زبان بھی بے نیاز نہیں ہے۔ مقتدرۂ قومی زبان اسلام آباد نے اِس سلسلے میں تگ و دو کی اور اردو ضابطہ تختی (code plate) کا پہلا ورژن سامنے آیا۔ ملاحظہ کیجیے: منسلکہ-۱

اردو ضابطہ تختی کے پہلے ورژن کے بعد کئی اجلاسوں میں اِس پر بحث مباحثہ ہوا اور بہتر سے بہتر کی طرف سفر شروع ہوا۔ اِس تفصیل اور ضابطہ تختی کی مختلف روایتوں (versions) کے لیے رک: http://www.nlauit.gov.pk/zabta.htm

اردو کے لیے یونی کوڈ کا نظام

یونی کوڈ کنسورشیم نے سب سے پہلے اپنے ورژن 3.01 میں اردو کے لیے ضمنی سہولت مہیا کی۔ دراصل دائیں سے بائیں لکھی جانے زبانوں میں سب سے پہلے عربی کے لیے یونی کوڈ میں سہولت دی گئی۔ لیکن حروفِ تہجی کی تعداد کے اعتبار سے چونکہ عربی کا ابجد فارسی کا تحتی

سیٹ ہے۔ اور فارسی کا ابجد اردو کا تحتی سیٹ بنتا ہے۔ اِسی لیے یہ نظام ذرا سی مشکل کا شکار ہوگیا۔ تاہم یہ حقیقت ظاہر ہونے کے بعد کہ اُردو میں الفاظ کی تعداد عربی اور فارسی سے زیادہ ہے، یونی کوڈ نے اردو کے لیے حروفِ تہجی داخل کرنے کا راستہ کھول دیا۔ ملاحظہ کیجیے: منسلکہ-۲

یونی کوڈ کے ورژن 4.0 اور 5.0 میں مزید حروف اور نشانات داخل کیے گئے۔ ملاحظہ کیجیے: منسلکہ-۳

اردو کے حروفِ تہجی میں حروف اور نقاط میں اِفتراق: پہلی کوشش

اوپر کی بحث میں یہ بات واضح ہوچکی ہے کہ عربی رسم الخط میں حروف پر نقطے اور اِعراب بہت بعد میں لگائے گئے۔ لیکن ٹائپنگ میں یہ ترتیب سب سے پہلے ۱۹۴۹ء میں محمد قطب الدین نے پیش کی۔ اُنھوں نے حروف کی قبیلہ وار تقسیم سے فائدہ اُٹھایا اور بتایا کہ اردو کے حروف بنیادی طور پر ۱۸ قبیلوں میں منقسم ہیں جنھیں محض چار طرح کے نقطوں اور چھوٹی ''ط'' کی مدد سے لکھا جاسکتا ہے۔ اِس تصور کی مکمل تفصیل ملاحظہ کیجیے: ۱۴

اردو کو مشین ریڈا یبل بنانے کے لیے محمد قطب الدین کے اِس تصور میں دورِ جدید کی ضروریات کے موافق بہت کچھ اضافے کرکے ڈاکٹر عطش درانی نے ۲۰۰۵ء میں کمپیوٹر اور اِنٹرنیٹ کے ماہرین سے اِستصواب کرکے اِسے یونی کوڈ کنسورشیم میں منظوری کے لیے پیش کیا۔ اگرچہ یونی کوڈ کنسورشیم اپنے ورژن 3.01 میں الفاظ کی خالی کشتیوں کو باقاعدہ نمبر تفویض کرچکا تھا ۱۵ لیکن نقطوں کو الگ حیثیت سے باقاعدہ نمبرات تفویض کرنے میں اُس کے تکنیکی ماہرین کو تحفظات تھے۔ اِس سے بھی بہت پہلے عربی میں ا، ٓ اور آ کے لیے الگ الگ نمبرات دیے گئے تھے۔ (اِن کے نمبر علی الترتیب یہ ہیں: ا=0627، ٓ=0653 اور آ=0622) جب کہ تکنیکی طور پر آ=ا+ٓ ہے۔ لیکن اِن سب حقائق کے باوجود اِس تجویز کی منظوری میں وقت لگ گیا۔ مقتدرۂ قومی زبان اسلام آباد کی طرف سے یونی کوڈ کنسورشیم سے اِس سلسلے میں کی گئی خط کتابت ملاحظہ کیجیے: منسلکہ-۴

یونی کوڈ کی نقطوں اور حروف کی خالی کشتیوں کے لیے جدا جدا نمبروں کی فراہمی

بالآخر یونی کوڈ کنسورشیم نے ۱۷/اپریل ۲۰۰۸ء کو یہ تجویز منظور کرکے ۲۲ تجاویز کو یونی کوڈ چارٹ میں جگہ تفویض کردی ہے۔ اِس کے مطابق اردو کے ۶، پشتو کے ۲، سندھی کے ۵ اور عربی کے ۱۰ نقاط اور مختلف شکلوں کو یہ جگہ ملی ہے۔ کوڈ چارٹ کے اندر یہ ۲۲ جگہیں 0800 تا 0889، 088A تا 088F اور 0890 تا 0895 تک دی گئی ہیں۔ اِس طرح اب حروف کی خالی کشتیاں اور مجرد نقطے الگ الگ حیثیت میں چہرہ دار ہوگئے ہیں۔ یونی کوڈ کے اِس تازہ چارٹ کی تصویر ملاحظہ کیجیے: منسلکہ-۵ (Arabic Glyph Parts)

لیکن یہاں اِس بات کی وضاحت ضروری ہے کہ یہ نقاط اور شکلیں صرف عربی کے لیے مخصوص نہیں ہیں بلکہ دائیں سے بائیں لکھی جانے والی کسی بھی زبان کے لیے مکمل طور پر کارگزار ہیں۔

نقطوں اور حروف کی خالی کشتیوں کے جدا جدا کردار کے فوائد

چونکہ اردو کا رسم الخط ایک معتدبہ حد تک عربی و فارسی کا پھیلاؤ (Extension) ہے اور یہ پاکستان میں بولی اور لکھی جانے والی قریب قریب تمام زبانوں کا یکساں رسم الخط بھی ہے، اِس لیے سب پاکستانی زبانیں اب اِسی ایک کوڈ چارٹ (0880) اور حروف کی خالی کشتیوں کی مدد سے لکھی جاسکتی ہیں۔ ایک نقطہ، دو نقطے، تین نقطے، چار نقطے، چھوٹی طوئے۔۔ نقطے اُلٹے چاہیے ہوں یا سیدھے وغیرہ، کا کسی بھی حرف کے اوپر یا نیچے لکھنا ممکن ہوگیا ہے۔ بلکہ نقطے کے اوپر چھوٹی طوئے اور حرف کے نیچے ش کے اوپر والے تین نقطے لگانا بھی ممکن ہے۔

نقطوں اور حروف کی جدا جدا حیثیت کی وجہ سے اردو بصری حرف شناسی (Optical Character Recognition) میں بھی بہت سہولت متوقع ہے۔

سب سے بڑی بات یہ ہے کہ صرف ۴۴ شکلوں کی مدد سے تمام پاکستانی زبانوں کا لکھا جانا ممکن ہوگیا ہے۔ بہت کم الفاظ کے استعمال کی وجہ سے Rendering میں بھی آسانی درآئی ہے۔

نقطوں اور حروف کی خالی کشتیوں کے الگ الگ کردار پر متوقع تحفظات

اِس ضمن میں ایک سوال یہ کیا جاسکتا ہے کہ کمپیوٹر کو حرف اور نقطے الگ الگ پراسیس کرنا پڑیں گے جس سے سپیڈ میں کمی ہوگی۔ دوسرا اِعتراض وہی ہوسکتا ہے جو محمد قطب الدین کے مجوزہ ٹائپ رائٹر پر کیا گیا تھا۔ کہ ٹائپسٹ کو ایک ہی وقت میں زیادہ کلیدیں دبانا پڑیں گی جس کی وجہ سے وقت زیادہ لگے گا۔

اِس سلسلے میں عرض ہے آج کے دور میں کمپیوٹر کی پراسیسنگ سپیڈ اِتنی ہوگئی ہے (اور ابھی تیز تر ہورہی ہے) کہ پراسیسنگ سپیڈ کا سوال ہی لایعنی ہوگیا ہے۔ دوسرے سوال کے دو جواب ہیں: آپ کوئی حرف جو مجرد شکل میں بھی آتا ہے اور نقطوں کے ساتھ کسی اور حرف کے طور پر بھی آتا ہے، جیسے مثلاً س جو تین نقطوں کے اضافے سے ش بن جاتا ہے۔ لکھیے تو یہ ازخود س+تین نقطے=ش بن کر treat ہوگا۔ اور اِس پراسیسنگ میں استعمال کنندہ کو کچھ نہیں کرنا پڑے گا۔ علاوہ ازیں یونی کوڈ کا ہر نیا چارٹ اپنے تمام پیش رو چارٹوں کو مکمل سہولت (support) دیتا ہے۔ لہٰذا یہ دونوں تحفظات بھی دور ہوگئے۔

ایک سوال redundancy کا بھی ہوسکتا ہے۔ اِس ضمن میں عرض ہے کہ یہ مسئلہ صرف یونی کوڈ کنسورشیم کا تکنیکی مسئلہ ہے۔ کمپیوٹر اور اِنٹرنیٹ کے کسی بھی استعمال کنندہ کو اِس سے کسی طور کوئی تنگی یا پریشانی نہیں ہوتی۔

خاتمہ

اردو کو آج کے مشینی دور میں کمپیوٹر اور اِنٹرنیٹ پر درست طور پر لکھنے، پڑھنے اور سمجھنے کے لیے اِس کا مکمل طور پر مشین ریڈایبل ہونا ضروری ہے۔ اردو حروف کو خالی کشتیوں اور نقطوں میں بانٹ کر کام کرنا وقت کی بہت اہم ضرورت ہے۔ یونی کوڈ کے 0880 Arabic Glyph Parts نے یہ ضرورت احسن طریقے پر پوری کردی ہے۔

حواشی:

۱۔ ''دِلّی کی آخری شمع-ش ح ح'' از حافظ صفوان محمد چوہان، مشمولہ جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان۔ شمارہ-۱۰ (۲۰۰۶ء)، ص-۱۹۳

۲۔ ''خط و خطاطی'' از شیخ ممتاز حسین جونپوری، اکیڈمی آف ایجوکیشنل ریسرچ، کراچی۔ ۱۹۶۱ء، ص-۳۱

۳۔ ''اردو رسم الخط اور ٹائپ'' از ڈاکٹر طارق عزیز، مقتدرۂ قومی زبان، اسلام آباد۔ ۱۹۸۷ء، ص-۱۴

۴۔ ''اردو رسمِ خط'' از محمد سجاد مرزا، عثمانیہ ٹریننگ کالج، حیدر آباد دکن۔ ۱۹۴۰ء، ص-۶

۵۔ بحوالہ ''اردو رسم الخط اور ٹائپ'' از ڈاکٹر طارق عزیز، مقتدرۂ قومی زبان، اسلام آباد۔ ۱۹۸۷ء، ص-۱۴

۶۔ ''اردو رسمِ خط'' از محمد سجاد مرزا، عثمانیہ ٹریننگ کالج، حیدر آباد دکن۔ ۱۹۴۰ء، ص-۶

۷۔ بحوالہ ''اردو رسم الخط اور ٹائپ'' از ڈاکٹر طارق عزیز، مقتدرۂ قومی زبان، اسلام آباد۔ ۱۹۸۷ء، ص-۱۶

۸۔ بحوالہ ''اردو املاء'' از رشید حسن خاں، نیشنل اکادمی، دریا گنج، نئی دہلی۔ ۱۹۷۴ء، ص-۴۸۳

۹۔ ''اردو رسم الخط اور ٹائپ'' از ڈاکٹر طارق عزیز، مقتدرۂ قومی زبان، اسلام آباد۔ ۱۹۸۷ء، ص-۱۹

۱۰۔ ایضاً، ص-۲۲

۱۱۔ ایضاً، ص-۲۳

۱۲۔ http://www.unicode.org/standard/WhatIsUnicode.html سے ترجمہ

۱۳۔ بحوالہ ''اردو رسم الخط اور ٹائپ'' از ڈاکٹر طارق عزیز، مقتدرۂ قومی زبان، اسلام آباد۔ ۱۹۸۷ء، ص-۲۷

۱۴۔ ''اردو رسم الخط اور ٹائپ'' از ڈاکٹر طارق عزیز، مقتدرۂ قومی زبان، اسلام آباد۔ ۱۹۸۷ء، ص-۱۷۴ تا ۱۷۹

۱۵۔ جیسے مثلاً ''ب'' کی خالی کشتی کے لیے 066E، ''ف'' کی خالی کشتی کے لیے 06A1 اور ''ق'' کی خالی کشتی کے لیے 066F۔ ملاحظہ کیجیے یونی کوڈ کا کوڈ چارٹ 0600-06FF

الف: منسلکات

۱۔ ''اردو ضابطہ تختی'' دوسرا ورژن
۲۔ یونی کوڈ 3.01 میں اردو کے لیے سہولت
۳۔ یونی کوڈ 4.0 اور 5.0 میں اردو کے لیے سہولیات
۴۔ یونی کوڈ کنسورشیم اور ڈاکٹر عطش درانی کی خط کتابت
۵۔ یونی کوڈ چارٹ: 0880 Arabic Glyph Parts

ب: اِنٹرنیٹ سائٹس (چند منتخب سائٹس)

1. http://www.microsoft.com

2. http://www.unicode.org

3. http://www.nla.gov.pk/beta/

ج: کتابیات

۱۔ رشید حسن خاں: ''اُردو املا''، نئی دہلی، نیشنل اکادمی، ۱۹۷۴ء۔

۲۔ طارق عزیز، ڈاکٹر: ''اُردو رسم الخط اور ٹائپ''، اسلام آباد مقتدرہ قومی زبان، ۱۹۸۷ء۔

۳۔ محمد سجاد مرزا: ''اُردو رسمِ خط''، حیدرآباد دکن، عثمانیہ ٹریننگ کالج، ۱۹۴۰ء۔

۴۔ ممتاز حسین جون پوری، شیخ: ''خط وخطاطی''، کراچی، اکیڈمی آف ایجوکیشنل ریسرچ، ۱۹۶۱ء۔

د: رسالہ

۱۔ ''جرنل ف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان''، شمارہ ۱۰، ۲۰۰۶ء۔

ہ: تکنیکی مشاورت

۱۔ ڈاکٹر خواجہ محمد زکریا، سابق پرنسپل، اورینٹل کالج، جامعۂ پنجاب، لاہور (لسانی اور اِصطلاحاتی تشکیلات وغیرہ کے لیے)

۲۔ ڈاکٹر عطش درانی، پراجیکٹ ڈائریکٹر، مرکزِ فضیلت برائے اردو اِطلاعیات، مقتدرۂ قومی زبان، اسلام آباد۔

و: مواد کی فراہمی میں معاونت

اِس مقالے کی تیاری کے دوران میں باربار جناب محبوب خان بگٹی، پروگرام منیجر اور محمد اکبر سجاد، اسسٹنٹ انفارمیٹکس آفیسر، مرکزِ فضیلت برائے اردو اطلاعیات، مقتدرۂ قومی زبان اسلام آباد کو زحمت دی گئی۔ میں اِن کا انتہائی شکر گزار ہوں۔

o ------- o

ضمیمہ

(I)

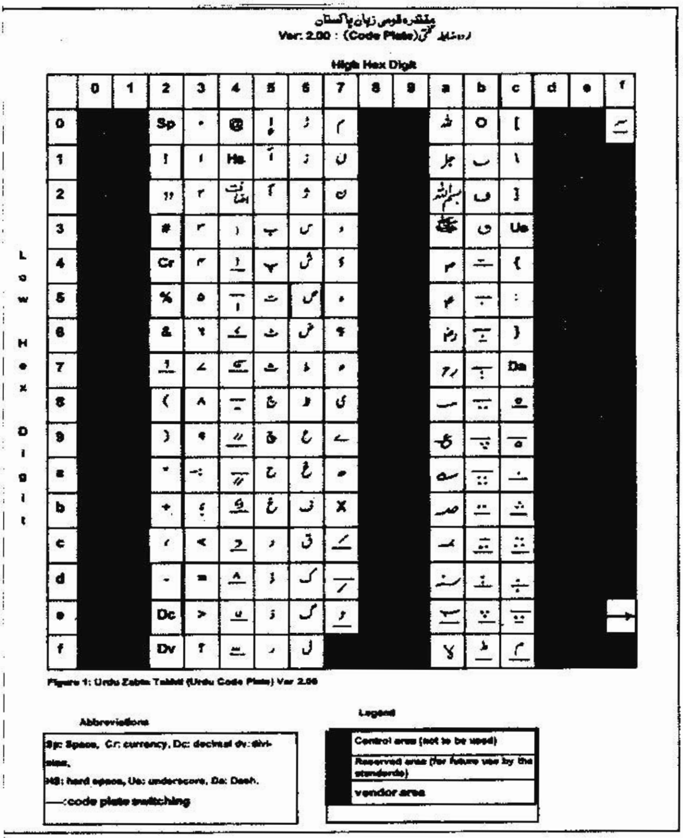
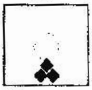
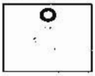
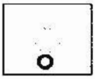
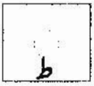


Figure 1: Urdu Zabta Takhti (Urdu Code Plate) Var 2.00

(Ii)

Character	Name	Code Point	Properties
	ARABIC SINGLE NUQTA ABOVE	*0604*	*Combining, Stacking*
	ARABIC SINGLE NUQTA BELOW	*0605*	*Combining, Stacking*
	ARABIC DOUBLE NUQTA ABOVE	*0606*	*Combining, Stacking*
	ARABIC DOUBLE NUQTA BELOW	*0607*	*Combining, Stacking*
	ARABIC TRIPPLE NUQTA ABOVE	*0608*	*Combining, Stacking*
	ARABIC TRIPPLE NUQTA BELOW	*0609*	*Combining, Stacking*
	ARABIC TRIPPLE INVERTED NUQTA ABOVE	*0616*	*Combining, Stacking*
	ARABIC TRIPPLE INVERTED NUQTA BELOW	*0617*	*Combining, Stacking*
	SINDHI QUADRPLE NUQTA ABOVE	*060A*	*Combining, Stacking*

(iii)

	SINDHI QUADRPLE NUQTA BELOW	*061A*	***Combining, Stacking***
	SINDHI DOUBLE DANDA ABOVE	*0618*	***Combining, Stacking***
	SINDHI DOUBLE DANDA BELOW	*0619*	***Combining, Stacking***
	SINDHI DOUBLE NUQTA VERTICAL ABOVE	*061C*	***Combining, Stacking***
	SINDHI DOUBLE NUQTA VERTICAL BELOW	*061D*	***Combining, Stacking***
	URDU SINGLE KASHIDA ABOVE	*063B*	***Combining, Stacking, Double Diacritic***
	URDU SINGLE KASHIDA BELOW	*063C*	***Combining, Stacking, Double Diacritic***
	URDU DOUBLE KASHIDA ABOVE	*063D*	***Combining, Stacking, Double Diacritic***
	URDU DOUBLE KASHIDA BELOW	*063E*	***Combining, Stacking, Double Diacritic***
	PASHTO SINGLE CIRCLE ABOVE	*063F*	***Combining, Stacking***
	PASHTO SINGLE CIRCLE BELOW	*065F*	***Combining, Stacking***
	URDU LETTER TOTA ABOVE		***Combining, Stacking***
	URDU LETTER TOTA BELOW		

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	؀	◌ؐ		ذ	ـ	◌ِ	٠	◌ٰ	ڀ	ڐ	ڠ	ڰ	ۀ	ې	◌۠	۰
1	؁	◌ؑ	ء	ر	ف	◌ّ	١	ٱ	ځ	ڑ	ڡ	ڱ	ہ	ۑ	◌ۡ	۱
2	؂	◌ؒ	آ	ز	ق	◌ْ	٢	ٲ	ڂ	ڒ	ڢ	ڲ	ۂ	ے	◌ۢ	۲
3	؃	◌ؓ	أ	س	ك	◌ٓ	٣	ٳ	ڃ	ړ	ڣ	ڳ	ۃ	ۓ	◌ۣ	۳
4		◌ؔ	ؤ	ش	ل	◌ٔ	٤	ٴ	ڄ	ڔ	ڤ	ڴ	ۄ	۔	◌ۤ	۴
5		◌ؕ	إ	ص	م	◌ٕ	٥	ٵ	څ	ڕ	ڥ	ڵ	ۅ	ە	ۥ	۵
6			ئ	ض	ن	◌ٖ	٦	ٶ	چ	ږ	ڦ	ڶ	ۆ	◌ۖ	ۦ	۶
7			ا	ط	ه	◌ٗ	٧	ٷ	ڇ	ڗ	ڧ	ڷ	ۇ	◌ۗ	◌ۧ	۷
8			ب	ظ	و	◌٘	٨	ٸ	ڈ	ژ	ڨ	ڸ	ۈ	◌ۘ	◌ۨ	۸
9			ة	ع	ى	◌ٙ	٩	ٹ	ډ	ڙ	ک	ڹ	ۉ	◌ۙ	۩	۹
A			ت	غ	ي	◌ٚ	٪	ٺ	ڊ	ښ	ڪ	ں	ۊ	◌ۚ	◌۪	ۺ
B	؋	؛	ث		◌ً	◌ٛ	٫	ٻ	ڋ	ڛ	ګ	ڻ	ۋ	◌ۛ	◌۫	ۻ
C	،		ج		◌ٌ	◌ٜ	٬	ټ	ڌ	ڜ	ڬ	ڼ	ی	◌ۜ	◌۬	ۼ
D	؍		ح		◌ٍ	◌ٝ	٭	ٽ	ڍ	ڝ	ڭ	ڽ	ۍ	۝	◌ۭ	۽
E	؎	؞	خ		◌َ	◌ٞ	ٮ	پ	ڎ	ڞ	ڮ	ھ	ێ	۞	ۮ	۾
F	؏	؟	د		◌ُ		ٯ	ٿ	ڏ	ڟ	گ	ڿ	ۏ	◌۟	ۯ	ۿ

The Unicode Standard 4.1, Copyright © 1991–2005, Unicode, Inc. All rights reserved.