Urdu Keyboards for Creating Bibliographic Records: An Assessment

N.S. Harinarayana
Associate Professor, Department of Studies in Library and Information Science, University of Mysore, Manasagangotri, Mysuru 570 006.
Email: ns.harianrayana@gmail.com

Arshiya Kouser
Project Assistant, Mysore University Library, University of Mysore, Manasagangotri, Mysuru 570 006. Email: arshiyalisc@gmail.com

Abstract

Input Method Editors (IMEs) – keyboard layouts - are the applications essential for converting keystrokes to character in another language. The study examined four different Urdu IMEs (keyboard layouts) to compare them for the creation of bibliographic records in libraries. It was found that there is no significant difference between the keyboards. The implication of the study is that the cataloguers can choose any one of the IMEs of their choice as there is no distinct advantage of any one IME over others.

Keywords

Input Method Editors, IMEs, Keyboards, Indic Keyboards, Urdu Keyboards, Perso-Arabic Scripts, Soft keyboards, Bibliographic Data Entry

Introduction

The development of Unicode has increased various language content in digital form. The databases, whose content was predominately in English earlier, also are being populated with content with multiple languages. The library databases are also becoming more multilingual, unlike the past. Libraries are encouraging, and users are demanding the metadata in the database to be in the language of the text of the catalogued item. The Unicode has solved the problem of encoding issues will be almost all the scripts that are known to the humanity so far. It is



likely that it remains the universal encoding standard for years to come.

The input of the non-Latin script-based languages is not as trivial as the Roman character based languages. The predominant input device for scripts remains to be the keyboard. The QWERTY keyboard is the most popularly available one. Applications have been developed to use the QWERTY keyboard for non-Latin-script languages also. Different techniques have been developed. While transliteration is one of the most common modes, the other approach is developing a language-specific input structure. INSCRIPT (INdian SCRIPT) keyboard layout was designed by Center for Development of Advanced Computing (CDAC) for Indic languages including Urdu. INSCRIPT, as a matter of fact, is a parallel development to ISCII (Indian Script Code for Information Interchange) – a character encoding standard for Indian languages that originate from Brahmi Script. The Bureau of Indian Standard (BIS) adopted the ISCII standard in 1991. It may be noted here that Unicode has retained the sequence of character codes in ISCII. The INSCRIPT keyboard was developed to enable the Indian scripts to be entered into the computer devices.

The present study aims to explore the efficiency of the freely available Input Method Editors (IMEs) for Urdu, which uses a Perso-Arabic script. Unlike other Indic languages, the Perso-Arabic scripts are rendered the right to left. As a matter of fact, Sindhi and Kashmiri also belong to the same category of language as that of Urdu as far as its script rendering is considered. Four IMEs were considered for testing purposes in this study.

Review of Literature

The most dominant input device to the computer is a keyboard. The QWERTY is popular among all keyboard layouts. Qwerty has its legacy in the age-old typewriters. The layout was designed to suit the mechanical tool - typewriters. But, even as the most sophisticated computers have come the same QWERTY layout has remained. Alternative keyboard structures have evolved before and after the invention of modern day computers. Still, the legacy of QWERTY cotinues. None of the other keyboard layouts have entirely replaced QWERTY and it has been the most widely used keyboard layout even today.

There are alternatives to QWERTY layout such as Dvorak, Coleman, Workman and so on. They have been compared for their efficiency by some studies in the past. The typing forces, finger movements, finger reach, muscle activity, comfort, etc.,

are some of the parameters used for comparing the keyboards. A consultancy company Cunningham & Cunningham, Inc.(1) gives a list of bad keyboards by highlighting the problems in the existing keyboard layouts.

Keyboard evaluation basically concentrates on testing the data entry rates. QWERTY was found to score over others in terms of efficiency (2,3) in a study which compared six keyboards: the QWERTY, ABC, Dvorak, Fitaly, JustType, and telephone. In this study the data entry rate was found to be 20.2 wpm for QWERTY which was significantly higher than the other layouts tested. In an interesting study, Norman (3), shows that alphabetically arranged keyboard layouts are slower than randomly arranged ones. He also shows that Dvorak's performance is only 5% better than the QWERTY keyboard which was originally designed by C.L. Sholes.

Soft keyboards, also called virtual keyboards or onscreen keyboards, are becoming popular with the advent of mobile and other handheld devices. The technology allows the soft keyboards to use different layouts including custom designed ones. Studies have been conducted to assess their performance. Mackenzie, Zhang & Soukoreff (4) developed a new ATOMIK layout after studying the existing Qwerty, FITALY, and OPTI designs. Again the rate of data entry has been the primary issue considered for design purpose. The size of the keyboards also changes in soft keyboards. In an interesting study (5) it is shown that size of the keyboards do not vary the speed of the data entry.

Indian efforts to design the keyboards go back to later part of 20th century. The keyboard layouts for most of the Indian languages were developed for typewriters. However, as far as the development of keyboard layout for computers is concerned, INSCRIPT by CDAC is the milestone. LIPIKAR, Talapatram, Panini, Google Indic, Quillpad, Keylekh, Sparsha to name a few are the keyboards for Indic languages. Mahesh and Sinha (6) give a comprehensive account on the Indian Scripts Processing activities. But, empirical study on the performance of the keyboards is hard to find in the literature. The present study attempts to provide empirical evidence on the suitability of keyboards based on the number of keystrokes required to enter the text.

Methodology

The present exploratory study is designed to test the efficiency of the IMEs for entering the bibliographic data in library databases. The IMEs considered in this

study are Madni Keyboard, Pak installer, Google input tools which are all based on the phonetic structure of keyboard, and Tahreer, which is INSCRIPT based keying developed by C-DAC.

The bibliographical details of ten books in the Urdu language available at Mysore University Library were selected for data entry purposes. The books were randomly selected. The particulars of the ten books were entered in Excel worksheet using all the four IMEs considered in this study. The keystrokes required for inputting the details were meticulously counted and tabulated for further analysis.

The keystrokes were grouped into two: simple and combined. The simple keystrokes involve the pressing of only one key at a time to get the desired display of Urdu script. Table 1 shows the word and the number of simple keystrokes required.

Word	Madni Keyboard	Pak installer	Google input tools	Tahreer INSCRIPT
Urdu	4	4	4	4
Zaban	4	4	4	4
Iqbal	5	5	5	5

Table 1: Simple Keystrokes

On the other hand, some of the letters in Urdu are displayed by pressing two or more keys at a time. These are considered grouped as 'combined keystrokes'. It is hypothesised in this study that the more the number of combined keystrokes required, the data entry will be slower. This has been tested in each of the data entry IME in the study.

As said earlier, the experiment was designed to test the number of keystrokes – simple and combined required for entering the bibliographic records using different IME's. The following table provided the one record of details of the data collected for the purpose.

The primary objective of the study is to help the regional cataloguers to make easy cataloguing by using right IME and by doing so we can save the time also.

There could be other methods for evaluation of IMEs. Time-motion study, finger movement study and so on. These kinds of studies are found for comparison of

QWERTY and Dvorak keyboards. But, the present study is simple and confined only to count the number of keystrokes required. The researchers agree that more comprehensive and more defined studies need to be carried out to establish the efficiency of the keyboard layout.

The characteristic features of each of the keyboards used in the study are given in the next section.

Urdu keyboards

The history of Urdu keyboard for a typewriter is more than a century-year-old. The first structure for keyboard in Urdu was evolved in the year 1911 (7). Like other keyboards, the Urdu keyboard also went through the stages metamorphosis before attaining the accepted structure as is used today. Modern improvements in Urdu keyboard were pioneered by the National Language Authority (Muqtadira-e-Qaumi Zaban) in Pakistan, which standardised the linguistic aspects such as orthography and lexicography (7). With the development of Information and Communication Technology (ICT), the keyboard structures have become compatible with computers regarding input of content/text. It appears that there is a dearth of systematic study on the efficiency of keyboards. The present study is a small step towards adopting a quantitative methodology for assessing the ability of keyboards and comparing them regarding keystrokes required to enter a corpus of text.

Urdu Phonetic Keyboard Layout

Urdu Phonetic Keyboard Layout was tested under the Windows environment in the present study. The underlying encoding standard for a character in all the keyboard structure is Unicode.

Madni Keyboard

Microsoft Windows provides Arabic / Urdu support, but Urdu Keyboard mapping for an average computer user is not available (8). Considering the shortfall of Windows default facility, Madni Keyboard was developed. Even Arabic language is also supported in Madni Keyboard. The layout of the keyboard is shown in figure 1. The major features of the keyboard are as follows:

- Arrangement of alphabets in the keyboard are phonetic
- It supports Urdu and Arabic languages.
- It is a Unicode supported keyboard layout.

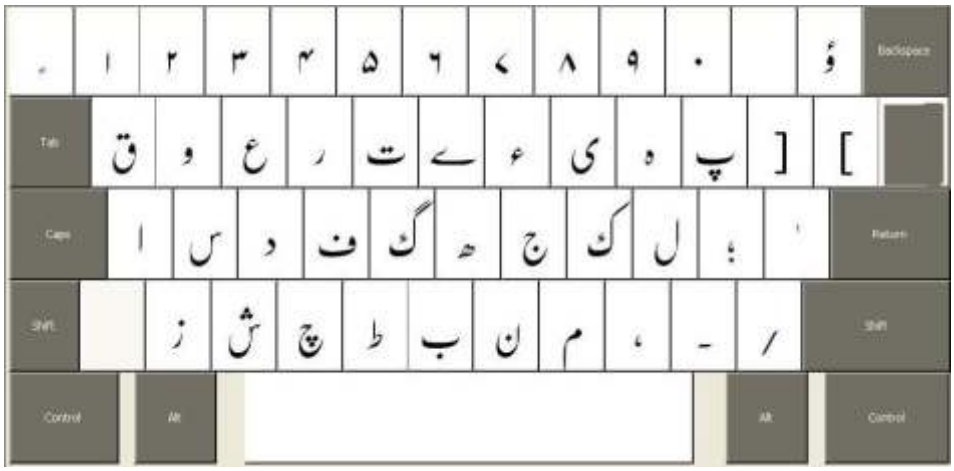


Figure 1 : Madni Keyboard Layout

Google Input Tool for Windows

Google Input Tools for Windows is an Input Method Editor (IME) which allows users to enter text in any of the supported languages using a Latin (English / QWERTY) keyboard. Users can type a word the way it sounds using Latin characters and Google Input Tools for Windows will convert the word to its native script. Available options in Google Input Tools for data entry purposes include transliteration, IME, and on-screen keyboards.

Google Input Tools for Windows is currently available in 22 different languages: Amharic, Arabic, Bengali, Persian, Greek, Gujarati, Hebrew, Hindi, Kannada, Malayalam, Marathi, Nepali, Oriya, Punjabi, Russian, Sanskrit, Serbian, Sinhala, Tamil, Telugu, Tigrinya and Urdu.



Figure 2 : Google Input Tools

Pak Urdu Installer

Developed basically for the Windows environment, the Pak Urdu Installer, referred as PUI in this article, is one of the famous Urdu IMEs for Urdu content creation in computers. It is available both in 32-bit and 64-bit versions. The PUI is entirely Unicode compatible. The PUI uses the standard QWERTY keyboard layout for inputting the Urdu script and was developed by PakSign Technologies (9).

The PUI leverages the basic facility to type and display Urdu characters already available in Windows environment. PUI elevates the quality of display of Urdu fonts. Urdu keyboard layout of PUI is required to be installed for typing the Urdu script. Once installed, the option for Urdu keyboard appears in the language bar of the computer.

It may be noted that the PUI is not a hardware solution like the GIST. PUI is an application that needs to be downloaded. On invoking the PUI, the standard keyboard acts like an Urdu keyboard. The keys are phonetic and arranged with corresponding similar sounding English alphabets, like "A" key is for "ALIF" and "B" key is for "BAY" etc. Windows provide an option to swap across other keyboard layouts which enable the users to type different languages.

PUI provides three fonts: artistic font "Jameel Noori Nastaleeq", CRULP font "Nafees Web Naskh" and BBC Urdu font "Urdu Naskh Asia Type". All these three fonts are Urdu Unicode True Type font (TTF).

Tahreer INSCRIPT layout keyboard

Two parallel developments took place in India in language computing during the late 1980s. CDAC provided the leadership in this area by developing ISCII and INSCRIPT Keyboard. Both these developments provided the boost for language computing efforts in India. As far as the INSCRIPT keyboard is concerned, the keyboard uses the standard QWERTY keyboard with 101 keys. One of the much-appreciated features of the INSCRIPT keyboard is that it maintains the same character mapping for all the languages of India that it supports. INSCRIPT keyboard layout, all the vowels are placed on the left side of the keyboard layout and the consonants, on the right side (10). The placement is such that the characters of one varg are split over two keys.

Tahreer is India's contribution through CDAC (11) for developing content in computers in Perso-Arabic Scripts - Urdu is one of them. The development of this

application has been enriched by C-DAC GIST's expertise in Indian Language Computing. CDAC has to come out with a separate effort for the Urdu language as it is a script written and read from right to left, unlike other Indian languages.

TAHREER is capable of handling Right to Left scripts in any UNICODE compliant application, such as - Notepad, MS-Word, etc. (in Windows 2000 and above editions with suitable Language Pack Installed). Using Tahreer along with any Unicode enabled web content creation tool, the user can create HTML documents which can be published on the Web.

The user can use the application for typing using various layouts (INSCRIPT/ PHONETIC/ CUSTOM) by selecting toggle key (Capslock / Numlock / ScrollLock). The tool also comes with a Virtual Onscreen Floating keyboard for easy typing using a mouse. The keyboard layout is shown in Figure 3. The salient features are:

•Support for typing Urdu, Sindhi and Kashmiri languages.

•Pleasing aesthetic NASKH and NASTAALIQ, UNICODE compatible Open Type fonts.

•The choice to type in Urdu, Sindhi & Kashmiri on any Unicode compatible Editor.

•User-friendly and straightforward graphical user interface.

•INSCRIPT keyboard layout is supported.

•PHONETIC keyboard is available for users who are more comfortable with English language typing.

•An Onscreen Virtual Floating Keyboard, to view the layout and to type in the corresponding Indian language.

•Choice of the toggle key / script key to change between languages is provided.

•Comparison of IMEs for Urdu script



Figure 3 : INSCRIPT keyboard layout for Urdu

The method used in the study for comparison of IMEs for Urdu script is simple and straight forward. The bibliographic text for sample entries was collected from the Urdu books available in Mysore University Library. The bibliographic details of randomly selected books in the Urdu language were collected and catalogued in Microsoft Excel sheet. The number of keystrokes required to enter the particulars of the books in all the four keyboards was meticulously counted and recorded in the spreadsheet. Table 1 shows the data collected for the study.

The simple keystrokes and combined keystrokes were recorded separately. Simple keystrokes are those which require only one key to be typed to form a character and those characters which need a combination of more than one keys for formation/display of a character. Table 2 shows the keyboard strokes recorded for entering the bibliographic details of ten books selected for the study.

Sl. No	Madni Keyboard			Google Input Tools			Pak Installer	Tahreer INSCRIPT keyboard				
1	71	6	77	72	5	77	70	6	76	71	5	76
2	154	5	159	157	2	159	153	4	157	147	9	156
3	101	7	108	101	7	108	99	9	108	99	7	106
4	92	5	97	95	2	97	91	5	96	89	7	96
5	100	6	106	100	6	106	99	6	105	95	10	105
6	50	8	58	53	5	58	52	7	59	52	5	57
7	160	12	172	161	11	172	164	7	171	165	7	172
8	160	8	168	163	5	168	159	7	166	157	9	166
9	89	4	93	91	2	93	88	4	92	89	2	91
10	107	8	115	108	7	115	106	8	114	108	6	114
Total			**1153**			**1153**			**1144**			**1139**

Table 2 : Keyboard Comparison Based on Keystrokes

As far as the number of keystrokes is concerned, there is no much difference between the strokes required by the keyboards. It may, probably, need more extensive testing by taking a large number of books. Even the ratio between the simple and combined keystrokes remain almost the same establishing the fact that there exists no difference between the keyboards as far as the number of keystrokes is concerned.

For the sake of academic discussion, one may note that the CDAC's Tahreer INSCRIPT Keyboard required the least number strokes (1139) for entering the sample text considered in the study. However, Google Input Tools uses a relatively lesser number of combined keystrokes in comparison with other keyboard layouts and thus probably reducing the speed of data entry. A more

comprehensive study is required to establish the efficacy of the keyboards on the speed of data entry and as such this study has not been designed to compare the IMEs regarding data entry speed.

The significant difference between Tahreer INSCRIPT Keyboard and the remaining keyboards is that the former is not based on phonetic transcription while the others are.

Table 3 examines the IMEs tested in this study for their consistency of keystrokes by using the coefficient of variance. The mean and the standard deviation are also shown in the table.

	Madini	Google Input Tools	Pak Installer	Tahreer INSCRIPT
Record	**Total Keystrokes**	**Total Keystrokes**	**Total Keystrokes**	**Total Keystrokes**
1	77	77	76	76
2	159	159	157	156
3	108	108	108	106
4	97	97	96	96
5	106	106	105	105
6	58	58	59	57
7	172	172	171	172
8	168	168	166	166
9	93	93	92	91
10	115	115	114	114
Mean	115.3	115.3	114.4	113.9
SD	36.91084	36.91084	36.3241	36.7735
CV	0.320129	0.320129	0.317518	0.322858

Table 3 : Variance of Keyboards

Table 3 clearly shows that there exists no significant difference between the keyboards in terms of their variations required for entering the texts. All of them have about 32% variations. This test also confirms that the cataloguers can choose any one of the keyboards for data entry purposes.

A few studies compare the English language keyboard layouts: QWERTY, Dvorak, Colemark, and Workman. Different methods have been used to compare them. In one of the methods, numerical scores have been assigned depending upon the finger movements from the home row of the key layout (12). Taking a clue from that study, similar method, if not the same one, is adopted to compare the IMEs used in this study.

For this study, each simple keystroke is multiplied with the score 1 and a combined keystroke with 0.75. For example, a bibliographic record requiring 30 keystrokes, would get the final value of 30 X 1 = 30. On the other hand, a bibliographic record which uses 10 combined keystrokes will have the final value of 10 X 0.75 = 7.5. This scoring is designed to indicate the speed with which the record can be entered. It is presumed that a combined keystroke takes more unit of time than the simple keystroke and hence gets slightly lesser score value. Table 4 indicates the overall result by adopting the above method.

Record	Madini Keyboard		Google Input Tools		Pak Installer		Tahreer INSCRIPT	
	Total Keystrokes	Product of total and score	Total Keystrokes	Product of total and score	Total Keystrokes	Product of total and score	Total Keystrokes	Product of total and score
1	77	75.5	77	73.5	76	74.5	76	74.75
2	159	157.75	159	158.5	157	156	156	153.75
3	108	106.25	108	106.25	108	105.75	106	104.25
4	97	95.75	97	96.5	96	94.75	96	94.25
5	106	104.5	106	104.5	105	103.5	105	102.5
6	58	56	58	56.75	59	57.25	57	55.75
7	172	169	172	169.25	171	169.25	172	170.25
8	168	166	168	166.75	166	164.25	166	163.75
9	93	92	93	92.5	92	91	91	90.5
10	115	113	115	113.25	114	112	114	112.5
Total	**1153**	**1135.75**	**1153**	**1137.75**	**1144**	**1128.25**	**1139**	**1122.25**

Table 4 : Keystrokes with Multiplying Factor

It is important to find out whether the difference between the total keystrokes of IMEs is statistically significant. Hence, an ANOVA test was performed on the data collected. The null hypothesis that all the IMEs work with same efficacy regarding the number of keystrokes required was formulated for the Anova test. The calculation needed to ANOVA test is shown in Table 5.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Madini	10	1135.75	113.575	1495.098
Google Input Tools	10	1137.75	113.775	1516.895
Pak Input Installer	10	1128.25	112.825	1465.612
Tahreer Inscript	10	1122.25	112.225	1477.02

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	15.225	3	5.075	0.003409	0.99972	2.866266
Within Groups	53591.63	36	1488.656			
Total	53606.85	39				

Table 5: Anova - Single Factor Calculation

Table 5 is the calculation of Anova for the four IMEs selected in the study. It is evident from the table that F value is lesser than the F-crit indicating the acceptance of the null hypothesis that all the IMEs used in the study are equal regarding the number of keystrokes required to enter the text.

Conclusion

Bibliographic records in library database are multilingual in nature. The choice of the keyboards affects the data entry work. There are a few studies on the comparison of standard English keyboards. Dvorak, for example, has been found better in performance than the most popular QWERTY keyboard. Such comparison studies of keyboard layouts have not been found in other languages, especially from libraries.

The present study undertook the comparison of designs of four IMEs, viz., Madini, Google Input Tools, Pak Installer and Tahreer Inscript. The catalogue data from Urdu books was taken to study these four keyboard layouts. The testing was done at three different levels. It was found out that as far as the Urdu keyboard layouts are concerned, all of them work in the same degree of efficiency as far as the number of keystrokes is concerned.

References

1. Bad Keyboards [Internet]. [cited 2016 Jul 7]. Available from: http://c2.com/cgi/wiki?BadKeyboards
2. Fadhilah MR, Santosa PI, Kusumawardani SS. Review Layout Keyboard Yang Optimal Untuk Pengetikan Dalam Bahasa Indonesia. SEMNASTEKNOMEDIA ONLINE. 2016 Feb 22;4(1):1-3–97.
3. Norman DA, Fisher D. Why Alphabetic Keyboards Are Not Easy to Use: Keyboard Layout Doesn't Much Matter. Hum Factors J Hum Factors Ergon Soc. 1982 Oct 1;24(5):509–19.
4. Mackenzie IS, Zhang SX, Soukoreff RW. Text entry using soft keyboards. Behav Inf Technol. 1999 Jan 1;18(4):235–44.
5. Sears A, Zha Y. Data Entry for Mobile Devices Using Soft Keyboards: Understanding the Effects of Keyboard Size and User Tasks. Int J Human–Computer Interact. 2003 Oct 1;16(2):163–84.
6. Sinha RMK. A Journey from Indian Scripts Processing to Indian Language Processing. IEEE Ann Hist Comput. 2009;31(1):8–31.
7. Urdu keyboard. In: Wikipedia, the free encyclopedia [Internet]. 2016 [cited 2016 Jul 8]. Available from: https://en.wikipedia.org/w/index.php?title=Urdu_keyboard&oldid=706118145
8. Urdu - Madni Keyboard [Internet]. [cited 2016 Jul 8]. Available from: http://urdu-madni-keyboard.software.informer.com/
9. Pak Urdu Installer - Urdu Fonts, Urdu Keyboard and Urdu Language Support Installer [Internet]. [cited 2016 Jul 8]. Available from: http://www.mbilalm.com/pak-urdu-installer.php
10. C-DAC: GIST - About GIST - INSCRIPT Keyboard Layout [Internet]. [cited 2016 Jul 7]. Available from: http://cdac.in/index.aspx?id=mlc_gist_inscript
11. C-DAC: GIST - Products - Tahreer [Internet]. [cited 2016 Jul 8]. Available from: http://www.cdac.in/index.aspx?id=mlc_gist_tahreer Dvorak keyboard typing vs. qwerty how do they compare? do these tests and see [Internet]. [cited 2016 Jul 7]. Available from: http://www.dvorak-keyboards.com/Dvorak_vs_qwertykeyboards.com/Dvorak_vs_qwerty_keyboard