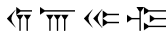


# Písma starého Orientu a moderní informační technologie

Otakar Smrž

Ústav formální a aplikované lingvistiky  
Univerzita Karlova v Praze

Katedra blízkovýchodních studií  
Západočeská univerzita v Plzni



# Přehled

- 1 Repräsentace
- 2 Arabština
- 3 Hebrejščina
- 4 Hieroglyfy
- 5 Klínopisy
- 6 Další písma
- 7 Appendix

# Informace, čísla, znaky

Počítače reprezentují informaci jako posloupnost **bitů**, tj. opozic právě dvou určitých hodnot. Tyto hodnoty lze chápat jako **symboly** dvojkové číselné soustavy. Série číslic se ztotožňují s **čísly**, která mohou odpovídat **písmenům** či **znakům** různých abeced, i tak mohou být běžně zobrazována.

Označení Unicode	Kódování UTF-8		Znak
Ux0639 0000 0110 0011 1001	0xD8 1101 1000	0xB9 1011 1001	arabsky ع
Ux004C 0000 0000 0100 1100	0x4C 0100 1100		latinkou L

# Informace, čísla, znaky

Počítače reprezentují informaci jako posloupnost **bitů**, tj. opozic právě dvou určitých hodnot. Tyto hodnoty lze chápat jako **symboly** dvojkové číselné soustavy. Série číslic se ztotožňují s **čísly**, která mohou odpovídat **písmenům** či **znakům** různých abeced, i tak mohou být běžně zobrazována.

Označení Unicode	Kódování UTF-8	Znak
Ux0639 0000 0110 0011 1001	0xD8 0xB9 1101 1000 1011 1001	arabsky ع
Ux004C 0000 0000 0100 1100	0x4C 0100 1100	latinkou L

# Informace, čísla, znaky

Počítače reprezentují informaci jako posloupnost **bitů**, tj. opozic právě dvou určitých hodnot. Tyto hodnoty lze chápat jako **symboly** dvojkové číselné soustavy. Série číslic se ztotožňují s **čísly**, která mohou odpovídat **písmenům** či **znakům** různých abeced, i tak mohou být běžně zobrazována.

Wikipedia > ASCII

<http://en.wikipedia.org/wiki/ASCII>

Wikipedia > UTF-8

<http://en.wikipedia.org/wiki/UTF-8>

# Unicode vlastními slovy

Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language.

Před vznikem Unicode existovaly stovky rozdílných kódovacích systémů pro přiřazování těchto čísel.

وَلَمْ يُوجَدِ نِظَامٌ تَشْفِيرٍ وَاحِدٌ يَحْتَوِي عَلَى جَمِيعِ الْمَحَارِفِ الصَّرُورِيَّةِ.

*Wa-lam yūğad nizāmu tašfirin wāhidun yaḥtawī ʿalā ġamāʿi 'l-maḥārifi 'd-ḍarūrīyati.*

Unicode

<http://www.unicode.org/>

# Unicode vlastními slovy

Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language.

Před vznikem Unicode existovaly stovky rozdílných kódovacích systémů pro přiřazování těchto čísel.

وَلَمْ يُوجَدَ نِظَامٌ تَشْفِيرٍ وَاحِدٌ يَحْتَوِي عَلَى جَمِيعِ الْمَحَارِفِ الصَّرْوِيَّةِ.

*Wa-lam yūğad nizāmu tašfirin wāḥidun yaḥtawī ʿalā ġamāʿi 'l-maḥāriḫi 'd-ḍarūrīyati.*

1608 1614 1604 1614 1605 32 1610 1615 1608 1580 1614 1583 32 1606 1616 1592 1614 1575 1605 1615 32 1578 1614 1588  
 1601 1616 1610 1585 1613 32 1608 1614 1575 1581 1616 1583 1612 32 1610 1614 1581 1578 1614 1608 1616 1610 32 1593  
 1614 1604 1614 1609 32 1580 1614 1605 1616 1610 1593 1616 32 1575 1604 1605 1614 1581 1614 1575 1585 1616 1601 1616  
 32 1575 1604 1590 1617 1614 1585 1615 1608 1585 1616 1610 1617 1614 1577 1616 46

# Unicode vlastními slovy

Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language.

Před vznikem Unicode existovaly stovky rozdílných kódovacích systémů pro přiřazování těchto čísel.

وَلَمْ يُوجَدِ نِظَامٌ تَشْفِيرٍ وَاحِدٌ يَحْتَوِي عَلَى جَمِيعِ الْمَحَارِفِ الصَّرُورِيَّةِ.

*Wa-lam yūǧad niḏāmu tašfīrin wāḥidun yaḥṭawī ʿalā ġamāʿi 'l-maḥārifi 'd-ḍarūrīyati.*

87 97 45 108 97 109 32 121 363 487 97 100 32 110 105 7827 257 109 117 32 116 97 353 102 299 114 105 110 32 119 257 7717  
 105 100 117 110 32 121 97 7717 116 97 119 299 32 703 97 108 257 32 487 97 109 299 703 105 32 700 108 45 109 97 7717 257  
 114 105 102 105 32 700 7693 45 7693 97 114 363 114 299 121 97 116 105 46



# Nepřímá transliterace

**Transliterace** je **přepis** jedněch symbolů jinými tak, aby zpětné nahrazení bylo možno provést **jednoznačně**.

**Unicode** je případ transliterace **přímé**, tzv. **one-to-one**. Notace **ArabTeXu** je proti tomu transliterace **nepřímá** neboli **meta-transliterace**.

وَلَمْ يُوَجِدْ نِظَامُ تَشْفِيرٍ وَاحِدٌ يَحْتَوِي عَلَى جَمِيعِ الْمَحَارِفِ الصَّرْوِيَّةِ.

*Wa-lam yūğad nizāmu tašfirin wāḥidun yaḥtawī ʿalā ġamāʿi 'l-maḥārifi 'd-ḍarūrīyati.*

## Romanization, Transcription and Transliteration

<http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/romanization.html>

# Nepřímá transliterace

Transliterace je **přepis** jedněch symbolů jinými tak, aby zpětné nahrazení bylo možno provést **jednoznačně**.

Unicode je případ transliterace **přímé**, tzv. **one-to-one**. Notace **ArabTeXu** je proti tomu transliterace **nepřímá** neboli **meta-transliterace**.

وَلَمْ يُوَجَدْ نِظَامٌ تَشْفِيرٍ وَاحِدٌ يَحْتَوِي عَلَى جَمِيعِ الْمَحَارِفِ الصَّرْوِيَّةِ.

*Wa-lam yūġad nizāmu tašfirin wāḥidun yaḥtawī ‘alā ġamī‘i ‘l-maḥārifi ‘d-darūrīyati.*

`\arabtrue \transfalse`

`\cap wa-lam yU^gad ni.zAmu ta^sfIriN wA.hiduN ya.htawI ‘alY  
^gamI‘i al-ma.hArifi a.d-.darUrIyaTi.`

# Nepřímá transliterace

Transliterace je **přepis** jedněch symbolů jinými tak, aby zpětné nahrazení bylo možno provést **jednoznačně**.

Unicode je případ transliterace **přímé**, tzv. **one-to-one**. Notace **ArabTeXu** je proti tomu transliterace **nepřímá** neboli **meta-transliterace**.

وَلَمْ يُوجَدَ نِظَامٌ تَشْفِيرٍ وَاحِدٌ يَحْتَوِي عَلَى جَمِيعِ الْمَحَارِفِ الصَّرْوِيَّةِ.

*Wa-lam yūġad niẓāmu tašfirin wāḥidun yaḥtawī ‘alā ġamā’i ‘l-maḥārifī ‘d-darūrīyati.*

```
\arabfalse \transtrue
```

```
\cap wa-lam yU^gad ni.zAmu ta^sfIriN wA.hiduN ya.htawI ‘alY  
^gamI‘i al-ma.hArifi a.d-.darUrIyaTi.
```

## Interpretace ArabTeXu

يُولَدُ جَمِيعُ النَّاسِ أَحْرَارًا مُتَسَاوِينَ فِي الْكِرَامَةِ وَالْحُقُوقِ. وَقَدْ وَهَبُوا عَقْلًا وَضَمِيرًا وَعَلَيْهِمْ  
أَنْ يَعامِلَ بَعْضُهُمْ بَعْضًا بِرُوحِ الإِخَاءِ.

```
\cap yUladu ^gamI'u an-nAsi 'a.hrAraN mutasAwIna fI al-karAmaTi  
wa-al-.huqUqi.
```

```
\cap wa-qad wuhibUA 'aqlaN wa-.damIraN wa-'alayhim 'an yu'Amila  
ba'.duhum ba'.daN bi-rU.hi al-'i_hA'i.
```

Interpretace ArabT<sub>E</sub>Xu

يولد جميع الناس أحرارا متساوين في الكرامة والحقوق. وقد وهبوا عقلا وضميرا وعليهم أن يعامل بعضهم بعضا بروح الإخاء.

```
\cap yUladu ^gamI'u an-nAsi 'a.hrAraN mutasAwIna fI al-karAmaTi
wa-al-.huqUqi.
```

```
\cap wa-qad wuhibUA 'aqlaN wa-.damIraN wa-'alayhim 'an yu'Amila
ba'.duhum ba'.daN bi-rU.hi al-'i_hA'i.
```

## Interpretace ArabTeXu

*Yūladu ġamīru 'n-nāsi ʾaḥrāran mutasāwīna fī 'l-karāmati wa-'l-ḥuqūqi. Wa-qad wuhibū ʿaqlan wa-ḍamīran wa-ʿalayhim ʾan yuʿāmila baḍduhum baḍdan bi-rūḥi 'l-ʾiḥāʾi.*

```
\cap yUladu ^gamI'u an-nAsi 'a.hrAraN mutasAwIna fI al-karAmaTi
wa-al-.huqUqi.
```

```
\cap wa-qad wuhibUA 'aqlaN wa-.damIraN wa-'alayhim 'an yu'Amila
ba'.duhum ba'.daN bi-rU.hi al-'i_hA'i.
```

Interpretace ArabT<sub>E</sub>Xu

يُولَدُ جَمِيعُ النَّاسِ أَحْرَارًا مُتَسَاوِينَ فِي الْكِرَامَةِ وَالْحَقُوقِ. وَقَدْ وَهَبُوا عَقْلًا وَضَمِيرًا وَعَلَيْهِمْ  
أَنْ يُعَامَلَ بَعْضُهُمْ بِبَعْضٍ بِرُوحِ الْإِخَاءِ.

يولد جميع الناس أحرارا متساوين في الكرامة والحقوق. وقد وهبوا عقلا وضميرا وعليهم  
أن يعامل بعضهم بعضا بروح الإخاء.

*Yūladu ġamīu 'n-nāsi aḥrāran mutasāwīna fī 'l-karāmati wa-'l-ḥuqūqi. Wa-qad wuhibū 'aqlan wa-ḍamīran wa-'alayhim 'an yu'amila baḍuhum baḍan bi-rūḥi 'l-iḥā'i.*

```
\cap yUladu ^gamI'u an-nAsi 'a.hrAraN mutasAwIna fI al-karAmaTi  
wa-al-.huqUqi.
```

```
\cap wa-qad wuhibUA 'aqlaN wa-.damIraN wa-'alayhim 'an yu'Amila  
ba'.duhum ba'.daN bi-rU.hi al-'i_hA'i.
```

## Buckwalterova transliterace

يُولَدُ جَمِيعُ النَّاسِ أَحْرَارًا مُتَسَاوِينَ فِي الْكِرَامَةِ وَالْحُقُوقِ. وَقَدْ وَهَبُوا عَقْلًا وَضَمِيرًا وَعَلَيْهِمْ  
أَنْ يُعَامَلَ بَعْضُهُمْ بِبَعْضٍ بِرُوحِ الْإِحَاءِ.

yuwladu jamiyEu {ln~aAsi OaHoraArFA mutasaAwiyna fiy  
{lokaraAmapi wa{loHuquwqi. waqado wuhibuWA EaqlAF  
waDamiyrFA waEalayohimo Oano yuEaAmila baEoDuhumo baEoDFA  
biruwHi {loIixaA'i.



## Buckwalterova transliterace

يولد جميع الناس أحرارا متساوين في الكرامة والحقوق. وقد وهبوا عقلا وضميرا وعليهم  
أن يعامل بعضهم بعضا بروح الإخاء.

ywld jmyE AlnAs OHrArA mtsAwyn fy AlkrAmp wAlHqwq. wqd  
whbwA EqLA wDmyrA wElyhm On yEAml bEDhm bEDA brwH AlIxA'.

## Buckwalterova transliterace

يُولَدُ جَمِيعُ النَّاسِ أحرَارًا مُتَسَاوِينَ فِي الكَرَامَةِ وَالْحُقُوقِ. وَقَدْ وَهَبُوا عَقْلاً وَضَمِيرًا وَعَلَيْهِمْ  
أَنْ يُعَامَلَ بَعْضُهُمْ بَعْضًا بِرُوحِ الإِخَاءِ.

yuwladu jamiyEu {ln~aAsi OaHoraArFA mutasaAwiyna fiy  
{lokaraAmapI wa{loHuquwqi. waqado wuhibuWA EaqlAF  
waDamiyrFA waEalayohimo Oano yuEaAmila baEoDuhumo baEoDFA  
biruwHi {loIixaA'i.

يولد جميع الناس أحرارا متساوين في الكرامة والحقوق. وقد وهبوا عقلا وضميرا وعليهم  
أن يعامل بعضهم بعضا بروح الإخاء.

ywld jmyE AlnAs OHrArA mtsAwyn fy AlkrAmp wAlHqwq. wqd  
whbWA EqLA wDmyrA wElyhm On yEAml bEDhm bEDA brwH AlIxA'.

# Encode::Arabic

- je **nezávislý** a **rozšiřitelný** nástroj pro zpracování elektronického textu
- umožňuje **konverzi** různých **reprezentací** arabštiny v několika **módech**
- lze využít ve významných **projektech** i při psaní běžných **dokumentů**

## Encode::Arabic Online Interface

<http://ufal.mff.cuni.cz/~smrz/Encode/Arabic/>

## Buckwalter Transliteration

<http://www.qamus.org/transliteration.htm>

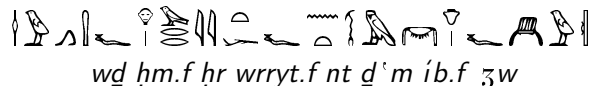
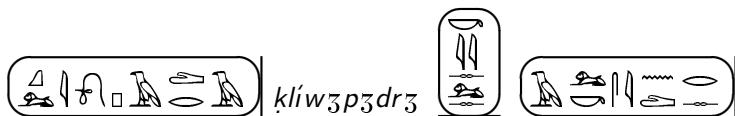
## ArabT<sub>E</sub>X User Manual

<http://129.69.218.213/arabtex/doc/arabdoc.pdf>

# Hebrejština v ArabT<sub>E</sub>Xu

Podle dokumentace je ArabT<sub>E</sub>X **typografický systém pro sazbu arabštiny i hebrejštiny**, autorovi prezentace se ale nepodařilo dokázat to na příkladu :)

# Egyptské hieroglyfy



L<sup>A</sup>T<sub>E</sub>X Archaic

<http://www.tug.org/tex-archive/fonts/archaic/>

HieroT<sub>E</sub>X

<http://www.tug.org/tex-archive/fonts/hieroglyph/>

# Egyptské hieroglyfy



*klíwzprz*



*wḏ ḥm.f ḥr wrryt.f nt ḏ'm íb.f 3w*

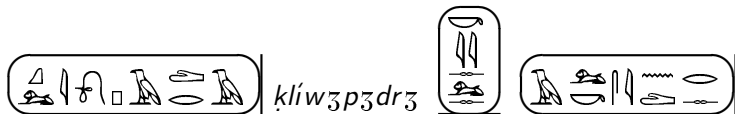
```
\Cartouche{\pmglyph{K:l-i-o-p-a-d:r-a}}
```

```
\translitpmhg{\HK\Hl\Hi\Ho\Hp\Ha\Hd\Hr\Ha}
```

```
\Vertouche{\pmvglyph{k:y:S:l:S}}
```

```
\Cartouche{\pmglyph{a-l:k-s-i-n:d-r:S}}
```

# Egyptské hieroglyfy



`\pmglyph`

`{\HU\Hw\HL\HJ\Hf-\Hq:\Hvbar-\HR:{\Hr:\Hr}-\Hy-\Ht:\HN-\Hf-\Hn:\Ht  
-\Hc\Hm\Hv-\HG:\Hvbar-\Hf\HZ\Hw\HV}`

`\translitpmhg`

`{\HU\ \HJ.\Hf\ \Hq\ \HR\Hr\Hy\Ht.\Hf\ \Hn\Ht\ \Hc\ \HG.\Hf\ \HZ}`







# L<sup>A</sup>T<sub>E</sub>Xovská rozšíření

ArabT<sub>E</sub>X moderní i historická arabština, perština, urdština, pašto, sindhi, ujugurština, malajština, hebrejština, jidiš

L<sup>A</sup>T<sub>E</sub>X Archaic písmo proto-semitské, fénické, řecké, etruské, runové, hieroglyfické, lineární, kyperské, ugaritské klínové, staroperské, aramejské, nabatejské, jihoarabské

HieroT<sub>E</sub>X

CJHebrew

FarsiT<sub>E</sub>X

Comprehensive T<sub>E</sub>X Archive Network

<http://www.ctan.org/>

# Documents as Data

document collection of data, logical rather than physical (cf. file)  
data information in the form suited for computer processing

We will explore how re-usable documents can be defined with the help of markup and a little bit of programming. We will separate the information content of our documents from their visual form. We will need  $\text{\LaTeX}$  and  $\text{\ArabTeX}$  to create one from the other.

Wikipedia >  $\text{\LaTeX}$

<http://en.wikipedia.org/wiki/LaTeX>

Wikipedia > XML

<http://en.wikipedia.org/wiki/XML>

# Expressing Meaning

Information requires **structure**. Explicit **annotation** of structure to a given **expression** in text or data is done through abstract **markup**. The **interpretation** of the markup in turn determines the **meaning** of the expression.

```
\begin{slide}
  \title{Expressing Meaning}

  Information requires \alert{structure}.
  Explicit \alert{annotation} of structure ...
\end{slide}
```

# Writing in L<sup>A</sup>T<sub>E</sub>X

Note the **notation** for the markup, esp. `\ { }`, and the **pairing** and possible **nesting** of `\begin{...}` and `\end{...}`.

```
\documentclass{article}
```

```
\begin{document}
```

This is the first paragraph of my document.

This is the second paragraph, completing the document for now. There is not much to it.

```
\end{document}
```

# Including ArabT<sub>E</sub>X

We can use ready-made **packages** to extend or change the **interpretation** of our document. We can include الحَطَّ العَرَبِيّ, of course.

```
\documentclass{article}
```

```
\usepackage{arabtex}
```

```
\begin{document}
```

```
    This is the first and only paragraph of my document
```

```
    \<wa-fIhA _ha.t.tuN 'arabIyuN> \RL{'aw h_aka_dA}.
```

```
\end{document}
```

# Flashcards!

`\documentclass{flashcards}` provides us with another functionality — it will typeset flashcards with vocabulary for us, with some **predefined** style, which we now modify explicitly to get the Arabic side both in the **orthography** and in the **transcription**.

```
\begin{flashcard}{sun beams}
  \RL{'asi' 'aTu as-samsi} % this text is a comment
  \\ \arabfalse\transtrue
  \RL{'asi' 'aTu as-samsi}
\end{flashcard}
```

```
\begin{flashcard}{official measures}
  \RL{'igrA'AtuN rasmIyaTuN} \\ \arabfalse\transtrue
  \RL{'igrA'AtuN rasmIyaTuN}
\end{flashcard}
```