Astar's guide into astroinformatics concepts

... or what I wish to knew when I was younger

Jaroslav Vážný

Masarykova univerzita

13.09 2011 / Vyškov

Abstract

Today's science lives in the virtual digitalized world. If we want to exploit the full potential of available data We have to be able to bump and grind in this world with ease and confidence. Computer science and technology in general are opportunities but require deep knowledge of the field. This is a problem because unlike mathematics computer science is not standard part of the scientific curriculum (at least not now in the Czech Republic). This presentation is meant to be short introduction in the important concepts in computer science. This is My personal point of view and it possible (and I hope) that other people see things in absolutely different light. What is my motivation? I have seen many brilliant physicists to struggle with simple tasks related with computers. I want to give young people some advices so they can deal with this subject with less pain.

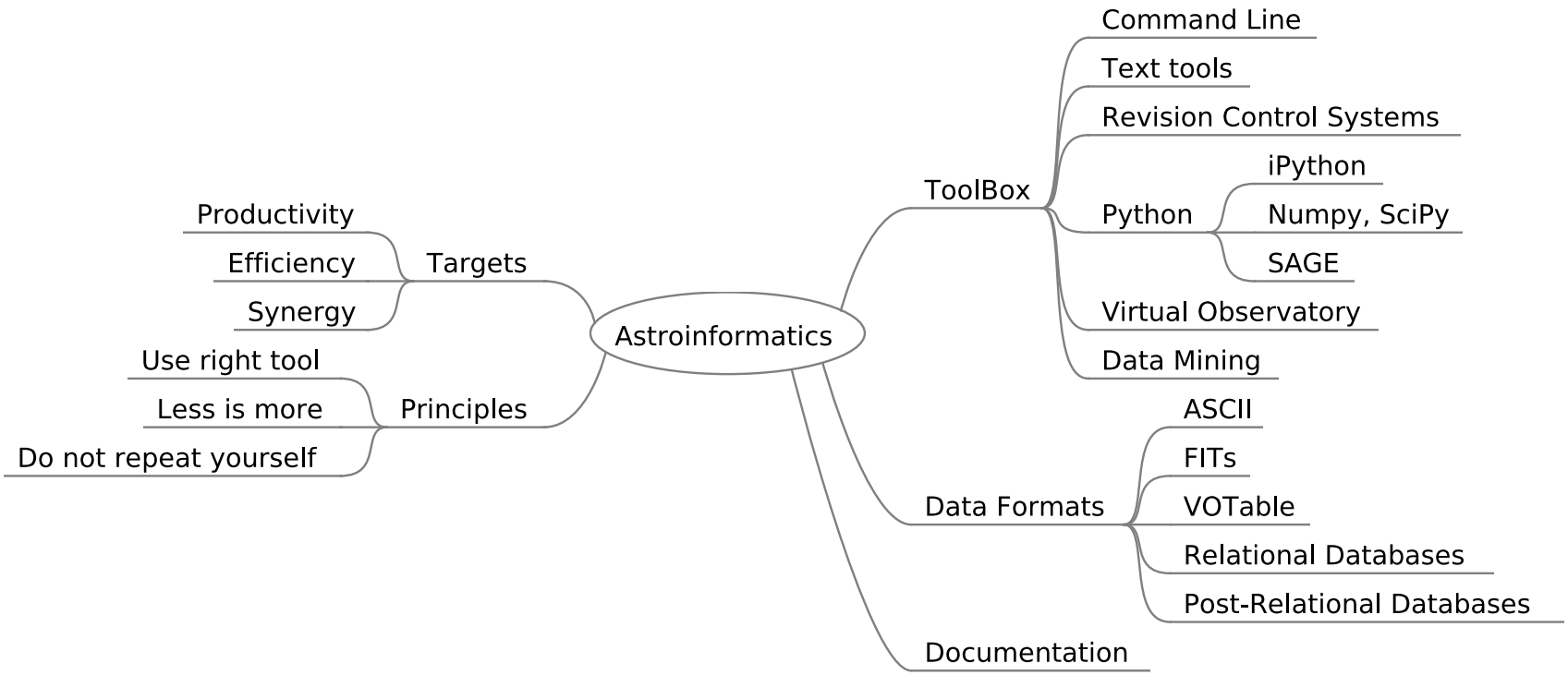
Motivation

There are two extreme cases one can see role of computers in science

- **Old school** Computer is just a tool. I want to focus on my science problem. I don't care what XML is.
- **Hackers** I want to know more ...;-)

  The best technique to avoid the troubles with computers is to have deep knowledge about wide concepts in computer science. Paradoxically both cases leads to the same conclusion!

Concepts introduced in this talk



Command Line

- Why is it important?
  - Efficient dialog computer $\Longleftrightarrow$ human
  - In all advanced tools (Programming, mathematica, CAD, . . .)
  - Cooperation, reusability, automatization
- Where I can learn it?
  - MUNI: PV004, F4270, PV065
  - PEEPCODE: Meet the Command Line, Advanced Command Line

Examples

- TAB
- !! Repeat last command
- !$ Repeat last agrument
- history command history
- CTRL+R search in history

Text tools

- Why is it important?
  - "Everything"is a text
  - head, tail, sed, awk, join, paste, vim, emacs ...
- Where I can learn it?
  - PEEPCODE: Meet Emacs, Smash Into Vim, Vim Emacs tutorials !!!

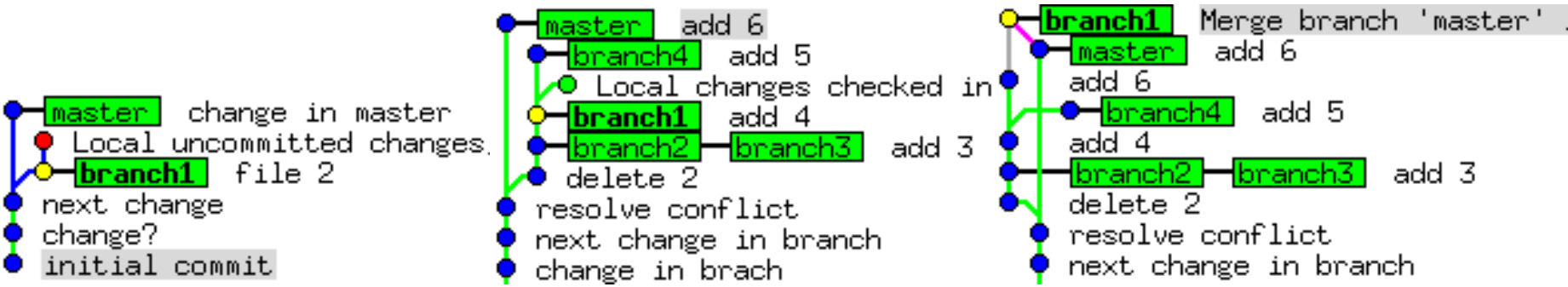
Examples

- echo "AlDeBaraN"| tr "[:upper:]""[:lower:]"
- vimdiff file1 file2

Revison Control Systems

- Why is it important?
  - Content history
  - Non-Linear Development
  - Cooperation
- Where I can learn it?
  - PEEPCODE: Git, Mercurial
  - `https://github.com`

Examples


master
change in master
Local uncommitted changes
branch1
file 2
next change
change?
initial commit
master
add 6
branch4
add 5
Local changes checked in
branch1
add 4
branch2
branch3
add 3
delete 2
resolve conflict
next change in branch
change in brach
branch1
Merge branch 'master'
master
add 6
add 6
branch4
add 5
add 4
branch2
branch3
add 3
delete 2
resolve conflict
next change in branch

Python

- Why is it important?
  - Real language (!= MATBLAB)
  - Easy to learn, Flexible
- Where I can learn it?
  - `http://www.archive.org`, Perez: Scientific Python, Google: Python in 2 days

iPython

- Why is it important?
  - Interactive Shell
  - Uniform access to system
  - New flexible design
- Where I can learn it?
  - `http://www.archive.org`, Perez: Scientific Python
  - `http://www.showmedo.org`

Example: Working with FITs in Python

```
In [1]: import atpy
In [2]: tbl = atpy.Table('spSpec-53401-2052-458.fit')
Auto-detected input type: fits
In [3]: tbl.write('votableExample.xml')
Auto-detected input type: vo
```

Updating FITS file.

```
In [1]: prihdr = hdulist[0].header
In [2]: prihdr.update('observer', 'Astar')
In [3]: prihdr.add_history('Updated 3/27/11')
```

Virtual Observatory

- Why is it important?
  - Uniform access to astronomy data
  - Based on Web standards
  - Nice GUI appliations;-)
- Where I can learn it?
  - `http://physics.muni.cz/~vazny/wiki/index.php/Diploma_work`

Example: Virtual Observatory Protocols

Cone Search Protocol

```
1 http://simbad.u-strasbg.fr/simbad-conesearch.pl?RA=24.5&
      DEC=-57.2&SR=0.1
```

Simple Image Access Protocol

```
1   http://hubblesite.org/cgi-bin/sia/hst_pr_sia.pl?POS
        =83.6,22.0&SIZE=1.0
```

Simple Spectra Access Protocol

```
1   http://archive.eso.org/apps/ssaserver/EsoProxySsap?
        REQUEST=queryData&POS=83.63,22&SIZE=1
```

Data Mining

- Why is it important?
  - Astrology of data
  - Data preprocessing
- Where I can learn it?
  - Standford(Andrew Ng)
  - `www.avc.cvut.cz`

Example: Decison Tree

```
ug <= 0.663668
|   gr <= -0.191208: 1 (7.0)
|   gr > -0.191208: 3 (104.0/5.0)
ug > 0.663668
|   ri <= 0.285854: 1 (88.0/5.0)
|   ri > 0.285854
|   |   ri <= 0.314657
|   |   |   gr <= 0.692108: 2 (6.0)
|   |   |   gr > 0.692108: 1 (3.0)
|   |   ri > 0.314657: 2 (90.0/2.0)
```

FITs

- Why is it important?
  - De-Facto standard in Astronomy
  - Flexible, Efficient, ASCII MetaData
- Where I can learn it?
  - `http://fits.gsfc.nasa.gov`

Example: Reading FITS file

```
In [1]: import pyfits
In [2]: hdulist = pyfits.open('spSpec-53237-1886-248.fit')
In [3]: hdulist.info()
Filename: spSpec-53237-1886-248.fit
No.    Name         Type      Cards   Dimensions   Format
0    PRIMARY    PrimaryHDU    213  (3874, 5)    float32
1               BinTableHDU    54  6R x 23C     [1E, 1E, ...
2               BinTableHDU    54  44R x 23C    [1E, 1E, ...
3               BinTableHDU    18  1R x 5C      [1E, 1E, ...
4               BinTableHDU    32  53R x 12C    [1J, 1J, ...
5               BinTableHDU    26  36R x 9C     [19A, 1E,
    ...
6               BinTableHDU    14  3874R x 3C   [1J, 1J, 1E]
```

Relational Databases

- Why is it important?
  - Sweetspot 100GiB – 1TiB
  - SQL = Efficient way to manipulate data
- Where I can learn it?
  - `http://www.sdss.org`

Example:Spectra from SEGUE project

```
1 SELECT objid,dbo.fGetUrlFitsSpectrum(s.specObjID)
2 FROM SpecPhotoAll s, platex p
3 WHERE s.specObjID is not null
4 AND s.plateid = p.plateid
5 AND p.programname LIKE 'segue%'
6 AND specClass = 1
```

VOTable

- Why is it important?
  - Standard in Virtual Observatory
  - Flexible, Efficient, XML
- Where I can learn it?
  - `http://www.ivoa.org`

Example: VOTable

```
<?xml version="1.0" encoding="utf-8"?>
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:noNamespaceSchemaLocation="http://www.ivoa.net/xml/
     VOTable/v1.0"
 xmlns="http://www.ivoa.net/xml/VOTable/v1.0">
 <RESOURCE type="results" >
  <TABLE >
   <FIELD ID="col0" name="wave" datatype="float" unit=""
   precision="F9"/>
  <DATA>
    <TABLEDATA>
     <TR>
      <TD>4012.50757</TD>
     </TR>
 </TABLEDATA>
   </DATA>
```

The power of TEX

- Why is it important?
  - Typography
  - Mathematics
  - Thesis, articles, presentations, posters, ...
- Where I can learn it?
  - MUNI: Plch, Sojka

Wake up!

Discussion