Perl for Linguists

Michael Hammond

University of Arizona
Perl for linguists
Perl for linguists

• Why programming?
Perl for linguists

- Why programming?
- Why Perl?
Perl for linguists

- Why programming?
- Why Perl?
- A current Perl project
Why programming?
Why programming?

• Collect data
Why programming?

• Collect data
• Analyze data
Why programming?

• Collect data
• Analyze data
• Remote computing
Why programming?

• Collect data
• Analyze data
• Remote computing
• Model theory
Why programming?

- Collect data
- Analyze data
- Remote computing
- Model theory
- General professional skills
Code

All the code for this presentation is available over the web (zipped) at the following URL:

http://linguistics.arizona.edu/~hammond/taiwan.html
Collecting data
Collecting data

• Running experiments locally (expprog.pl)
Collecting data

- Running experiments locally (`expprog.pl`)
- Running experiments locally with a GUI (`tkexp.pl`)
Collecting data

- Running experiments locally (`expprog.pl`)
- Running experiments locally with a GUI (`tkexp.pl`)
- Assembling corpora from local static resources (`makecorpus.pl`)

Analyzing data
Analyzing data

- Looking for patterns (*visgrep.pl*)
Analyzing data

- Looking for patterns (visgrep.pl)
- Counting things (neightk.pl)
Analyzing data

- Looking for patterns (visgrep.pl)
- Counting things (neightk.pl)
- Finding verbs (verbs.pl)
Remote computing
Remote computing

- Running experiments remotely (Bailey & Hahn replication, bhrep.cgi)
Remote computing

- Running experiments remotely ([Bailey & Hahn replication](#), bhrep.cgi)
- Interacting with local or remote databases (generating sql: dbiex.pl)
Remote computing

- Running experiments remotely (Bailey & Hahn replication, bhrep.cgi)
- Interacting with local or remote databases (generating sql: dbiex.pl)
- Assembling corpora from the web (websearch.pl)
Modeling theory
Modeling theory

• Optimality Theory (web interface, sylpars.pl)
Modeling theory

• Optimality Theory (web interface, sylpars.pl)

• N-gram models (a bunch of examples from a course on Statistical NLP that I did recently)
General professional skills
General professional skills

• General programming skills
General professional skills

- General programming skills
- Web programming
Why Perl?
Why Perl?

- Free
Why Perl?

- Free
- Multi-platform
Why Perl?

• Free
• Multi-platform
• Easy
Why Perl?

- Free
- Multi-platform
- Easy
- Multiple dialects
Why Perl?

- Free
- Multi-platform
- Easy
- Multiple dialects
- Powerful regular expression tools
Why Perl?

• Free
• Multi-platform
• Easy
• Multiple dialects
• Powerful regular expression tools
• Written by a "linguist"
Why Perl?

- Free
- Multi-platform
- Easy
- Multiple dialects
- Powerful regular expression tools
- Written by a “linguist”
- Perl poetry
Why Perl?

• Free
• Multi-platform
• Easy
• Multiple dialects
• Powerful regular expression tools
• Written by a “linguist”
• Perl poetry
• Obfuscated perl, “japhs”, etc.
Where to find out more
Where to find out more

- In any perl implementation the `perldoc` command can be used to find out lots and lots of stuff.
Where to find out more

• In any perl implementation the **perldoc** command can be used to find out lots and lots of stuff.

• The official and best perl website is [www.cpan.org](http://www.cpan.org), but see also [www.perl.org](http://www.perl.org).
Where to find out more

• In any perl implementation the `perldoc` command can be used to find out lots and lots of stuff.

• The official and best perl website is [www.cpan.org](http://www.cpan.org), but see also [www.perl.org](http://www.perl.org).

• ActiveState Perl (a free Windows implementation that I use) comes with extensive web-based documentation.
A current project
A current project

• “Experimental Syntax Server” (with Wayne Cowart, James Myers, and Keith Alcott)
A current project

- “Experimental Syntax Server” (with Wayne Cowart, James Myers, and Keith Alcott)
- Goal: to allow researchers to design, construct, run, and analyze simple experiments over the web.
A current project

• “Experimental Syntax Server” (with Wayne Cowart, James Myers, and Keith Alcott)

• Goal: to allow researchers to design, construct, run, and analyze simple experiments over the web.

• Today: a very simple outline of how such a system might work.
General logic of project
General logic of project

- Perl program running on a internet-accessible server.
General logic of project

- Perl program running on a internet-accessible server.
- Users interact via web-based forms.
General logic of project

- Perl program running on an internet-accessible server.
- Users interact via web-based forms.
- Program allows for experiments to be constructed, run, and have the results analyzed.
General logic of project

• Perl program running on a internet-accessible server.

• Users interact via web-based forms.

• Program allows for experiments to be constructed, run, and have the results analyzed.

• The construction and analysis functions make use of a graphical user interface and are simple enough for the naive user.
Some preliminary snapshots
Some preliminary snapshots

- Login screen
Some preliminary snapshots

• Login screen
• Not a registered user
Some preliminary snapshots

- Login screen
- Not a registered user
- A registered user
Some preliminary snapshots

- Login screen
- Not a registered user
- A registered user
- Choose a task
Some preliminary snapshots

- Login screen
- Not a registered user
- A registered user
- Choose a task
- Not yet implemented
Some preliminary snapshots

• Login screen
• Not a registered user
• A registered user
• Choose a task
• Not yet implemented
• Current experiments
Some preliminary snapshots

- Login screen
- Not a registered user
- A registered user
- Choose a task
- Not yet implemented
- Current experiments
- A sample experiment
The big picture

user
webbrowser

server
webserver
CGI program

other server
MySQL
MySQL database
MySQL database

• MySQL is a particular open-source database server.
MySQL database

- MySQL is a particular open-source database server.
- For the ESS project, it holds the following data:
MySQL database

- MySQL is a particular open-source database server.
- For the ESS project, it holds the following data:
  - Users and passwords
  - Experiments
  - Experimental items
  - Results from individual experimental runs
Programming challenges for ESS project
Programming challenges for ESS project

• The program runs on a separate machine.
Programming challenges for ESS project

- The program runs on a separate machine.
- The program makes use of several special perl modules:
  - CGI.pm: interact with the user;
  - DBI.pm: interact with MySQL.
Programming challenges for ESS project

- The program runs on a separate machine.
- The program makes use of several special perl modules:
  - CGI.pm: interact with the user;
  - DBI.pm: interact with MySQL.
- The program includes bits in HTML to create web pages.
Programming challenges for ESS project

- The program runs on a separate machine.
- The program makes use of several special perl modules:
  - CGI.pm: interact with the user;
  - DBI.pm: interact with MySQL.
- The program includes bits in HTML to create web pages.
- The program includes bits in SQL to talk to the database server.
Conclusions
Conclusions

1. Why programming
Conclusions

1. Why programming
2. Why perl
Conclusions

1. Why programming
2. Why perl
3. Where to find out more
Conclusions

1. Why programming
2. Why perl
3. Where to find out more
4. “Experimental Syntax Server” prototype