



COMP2720: Automating Tools for New Media

Guest lecture: Python in the
real world



Lecture Outline

- Something about me (Peter Christen)
- Why I like Python and use it every day
- The use of Python in my research
- Python in the real world – Some success stories
 - Some quotes from organisations that use Python
 - Google
 - Industrial Light and Magic (Star Wars)
 - United Space Alliance (NASA contractor)
 - Super League (UK's premier rugby league)
- Conclusions

Something about me

- Senior lecturer in Department of Computer Science
 - Been at DCS since 2001, before that Post-Doc in RSISE Computer Sciences Lab.
 - PhD in Switzerland (1999), where I also did my undergraduate, high-school, primary school, kindergarten, and where I was born.
- I have been using Python since 2000
 - For my research in data mining and data matching
- I originally developed and ran COMP2720 in 2005 and 2006
 - So you can blame me, not Alexei, for the stuff you don't like ;-)
 - But blame him for assignments and the exam ;-)
- More: <http://cs.anu.edu.au/people/Peter.Christen>



Why I like Python and use it every day

- I always have a Python terminal handy
 - Use it as a simple calculator
 - Or run short sequences of commands
- Over the years I have written many helpful scripts
 - For example, check word repeats in in text documents
 - Run image processing on all files in a directory (use Linux 'convert' program), scale and sharpen images for a Web site, and rename them
 - Run customised back-up jobs (copy my current project files onto a backup machine or a USB drive – but only certain file types)
- I have used Python for my research since 2000



Use of Python in my research

- My research interests are in data mining and data matching
 - **Data mining:** Analysis of very large databases, with the aim to find novel, useful and interesting patterns and rules.
For example: predict credit card fraud, predict how best to place items in a supermarket to increase sales, segment customer groups for targeted advertisement, etc.
 - **Data matching:** Find records from several databases that refer to the same person (customer, patient, etc).
For example to analyse effects of car accidents on health system (need data from hospitals, police, car insurances, health insurances, etc.)
- Research is about developing new algorithms, methodologies and techniques (prototypes), and Python is an ideal language for this



Use of Python in my research (2)

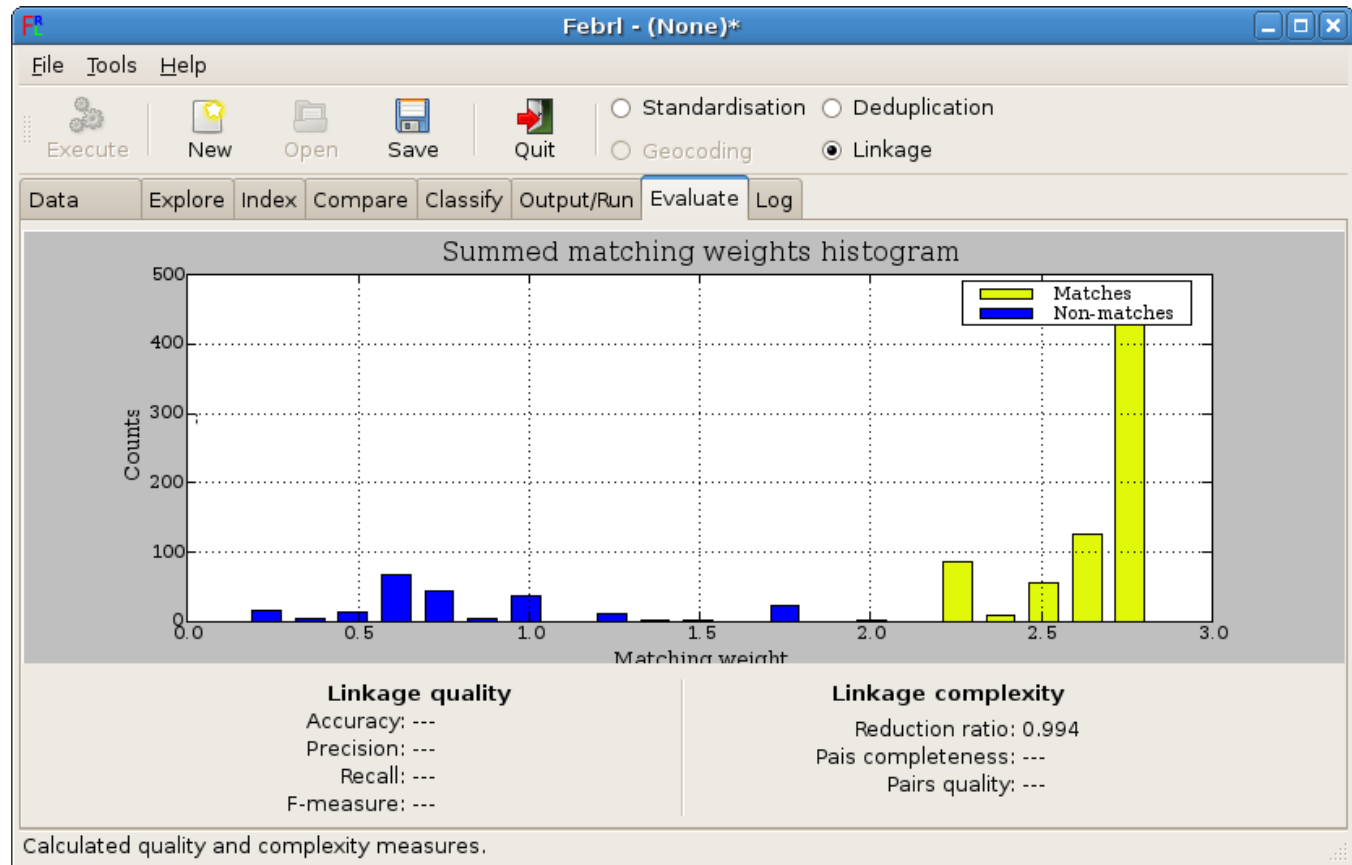
- Rapid prototype development with Python
 - Often, an idea can easily be written in Python because it provides a very large number of modules with diverse functionalities
 - No need to worry about low level programming issues (memory allocations, pointers, input/output, etc.)
 - Much faster to program an idea – thus much more productive
 - Easy to read, as there is no cluttering code required, even after months or years (compared to C / Java)
 - Performance (speed) is often not that important for a research prototype (unless your research is in performance and speed, like computer systems)
 - Easy to test (unit testing) and profile (find timing bottlenecks), and also easy to make faster (for example include C code)

Use of Python in my research – Example

- A core component of data matching is approximate string comparisons
 - Names and addresses in real-world databases often have typographical errors and variations, for example:
 - *Dr Peter Christen, Dept Computer Science, A.N.U.*
 - *Christen, Peter; ANU Department of Computer Sciences*
 - *Mr P. Christensen, computer Science, Austr. Natl. University*
 - Many different techniques on how to approximately compare strings (aim is to get a similarity value: 0=not similar at all, 1=exact match)
 - Python's many string functionalities allowed easy and quick development of a library of string comparison functions
 - No need to worry about string lengths, buffer memory allocations, etc.

Use of Python in my research – Example (2)

- I have developed a full data matching system using only Python





Python in the real world – Success stories

- Python is being used by many organisations, for examples see: <http://www.python.org/about/success>
- Application domains include:
 - Software development
 - Arts
 - Business
 - Education
 - Governments
 - Scientific
 - Engineering

Quotes about Python

- Taken from: <http://www.python.org/about/quotes>
 - "Python is fast enough for our site and allows us to produce maintainable features in record times, with a minimum of developers," said Cuong Do, Software Architect, *YouTube.com*.
 - "Python has been an important part of Google since the beginning, and remains so as the system grows and evolves. Today dozens of Google engineers use Python, and we're looking for more people with skills in this language." said Peter Norvig, director of search quality at *Google, Inc.*
 - "Python is everywhere at ILM. It's used to extend the capabilities of our applications, as well as providing the glue between them. Every CG image we create has involved Python somewhere in the process," said Philip Peterson, Principal Engineer, Research & Development, *Industrial Light & Magic*.



Python at Google

- Python is one of three official languages used by Google (others are C++ and Java)
- Python is used for its adaptability, rapid development and because it is easy to maintain
- Python is used at Google for:
 - Reporting (analysing logs and create reports)
 - Monitoring, restarting and data collection of production servers
 - Packaging of other codes so they can easily be distributed
 - Google build system (check dependencies, style, etc. of other code)
- Python programs at Google must follow strict style guidelines and extensive testing (incl. 2 people looking at it)



Python at Industrial Light and Magic (ILM)

- ILM was started in 1975 by George Lucas (Star Wars)
 - First computer generated movie sequence in 1982 (Star Trek)
- Rendering computer graphics in movie quality requires a lot of computational power (large number of workstations)
- Python used in ILM:
 - Can be embedded into larger applications
 - Python can invoke specific functions in a larger system
 - Python wrappers written around many ILM custom applications (similar to Visual Python)
 - Python is also used to develop a CG artist's interface to the ILM system



Python at United Space Alliance

- Main contractor for NASA's space shuttle program
- Requires quality just-in-time engineering solutions at low costs
- Python has been used since 1994 for:
 - Workflow Automation System (Python manages interactions between tasks, connects to Oracle database, and the graphical user interface is written in Python)
 - Python is advantageous because of easy code maintenance (easy to read even without extensive comments compared to C/Java)
 - Even non-programming engineers learned Python quickly



Python at Super League

- The UK's premier Rugby League
- Python used for Web site development of Super League (including news, e-mails, match and ticket management)
 - Pollenation Internet (Internet provider) approached as developer of Super league's Web site
 - Pollenation's customer's don't really care what language used for a project (but Pollenation wants to streamline their development)
 - Use of Python Web frameworks (Nevow and Twisted) and database support allowed fast development and in-time delivery
 - System is able to serve 300 requests per second (on 1 GHz server with 1 Gigabytes of main memory) in mid 2005



Conclusions

- Python is a language being used by many organisations in the real world
 - Good to know how to program Python, good to have on your CV
 - Quick search at <http://it.seek.com.au> for 'python': 175 matches, 9 in the ACT (on 18 October)
- Also a good language to do research in (rapid prototype development)
- Think about using Python for small specific tasks