

WebOOB (Web Outside Of Browser)

Use the Web from your shell!

François Revol
revol@free.fr



Who needs browsers anyway?

- “The Web is about transmitting **information** to everyone regardless the platform” (Tim B. Lee)
- Browsers need to load 2 or 3MB of images and JS even when you just need the data itself.
 - JS runs untrusted non-free code on your machine
- You can't easily pipe the browser into grep, cut or sed 😞



WebOOB, a Web client for your shell

- A python framework for web scraping
 - Several capabilities (video, bank, message...)
- CLI & GUI **applications** using capabilities
 - To search and collect data, submit forms...
- **Modules** implementing some capabilities
 - Youtube, Europarl (video), PhpBB (message)...



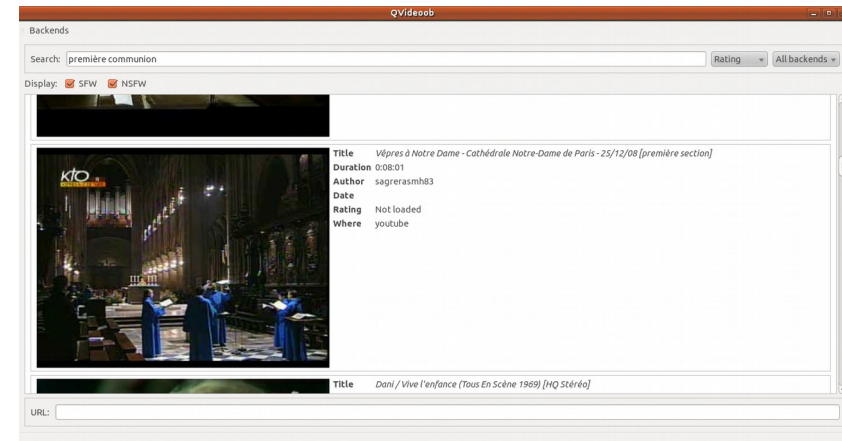
WebOOB framework

- A set of python classes
- Browser functions
 - HTTP[S] engine
 - HTML parser...
- Settings for application and backends
- Module discovery...



Applications

- Command line
 - Interactive (FTP-like commands) or not
 - Formatters for CSV, JSON, HTML, plain text...
- GUI (PyQt)
 - Simple GUI for a single task
 - “There's an OOB for that!”



(Some) Applications

- [Q]Boobmsg
- [Q]Cineoob
- [Q]Cookboob
- [Q]HaveDate
- [Q]Videoob
- Boobank
- Boobill
- Boobtracker
- Comparoob
- Pastroob...



Modules

- Support one or more capabilities for a website
- Instantiated for a specific website = *backend*
 - [vimeo]
 - _enabled = 1
 - _module = vimeo
 - [redminedemo]
 - _module = redmine
 - url = <http://demo.redmine.org/>
 - username = import



(Some) Modules

- Redmine
- Github (tickets)
- FreeMobile (bills)
- Many (french) banks
- Chronopost
- Collissimo...
- Youtube
- Europarl (videos)
- Vimeo
- Dailymotion...
- RMLL \o/ (videos)



Development status

- Not all modules support all wanted capabilities
 - Some video modules lack search function...
- Browser2 class makes writing modules easier
 - Some still needs rewriting from old Browser class
- Used professionally for banking websites



*nix commands composition

- Now you can
 - Redirect stdout to the Web
 - Redirect stdin from it as well
 - Automate things with your shell of choice
 - Support new sites without changing the workflow



Creating 200 tickets from a CSV?

- Configure a backend with the redmine or github account
- Parse the CSV, generate an mbox-like file / line
 - Properties as headers
 - Description as body
- ```
for f in *.txt; do boobtracker -d
post $account < $f; done
```
- Profit!



# Converting forum posts to slides

- `boobmsg -q -b phpbb formatter json ';' export_thread 36.1681 > talks.json`
- Some python to generate html slide templates:  
`python gen-desc.py talks.json`
- Convert them to PDF: `lowriter talks.html`



# Forum posts to slides

Applications Raccourcis 1 °C lun. 6 juil., 23:32 François Revol

slides\_stands.pdf - Description des stands Alchimie X

4 sur 17 70,71%

Vignettes

1

2


3

4

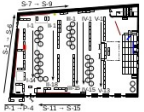


5


6

7

**[S-5] Planète Robots** 


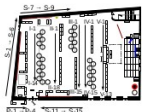
Planète Robots est le seul magazine dédié à la robotique, en langue française, dans le monde, et ce depuis décembre 2009. Le magazine propose également une série de services par le biais de son site internet comme la possibilité de déposer son CV sur la plateforme interne Jobs, dédiée à l'univers de la robotique. <http://www.planeterobots.com>




**[S-10] Fontaine "Data Flow" (art conceptuel)** 


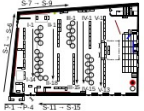
Cette installation artistique est une allégorie de la circulation des données, leur passage dans les "tuyaux" des réseaux et sur les supports physiques, et la vanité des tentatives de filtrage de celles-ci...


Concepteur : mmu\_man



**[S-10] Jeux de plateau** 

G2L2 Corp vous proposent de jouer à des jeux de plateau dans-la-vie-vraie, parce qu'il n'y a pas que des ordis dans la vie...





**[S-11] G3L : GNU/Linux et Logiciels Libres en Drôme-Ardèche** 

G3L est le Groupe d'Utilisateur de GNU/Linux et des Logiciels Libres (GULL) de Drôme-Ardèche.

Les réunions ont lieu deux vendredis par mois à la MJC Chateaufort de Valence.

<http://g3l.org/>



2015 [revol...] [icons] [[nvda...] [[slee... [Images] Table... [revol...] [revol...] [qemu...] [revol...] RMLL2... [revol...] stands... slides... slides...




# Forum posts to slides

Applications Raccourcis 1 °C lun. 6 juil., 23:32 slides\_conf.pdf — Conférences Alchimie X

6 sur 14 71,90%

Vignettes

**[Sam 18:00] Vente de produits régionaux** 

Des producteurs locaux vous proposent de nombreux produits non-numériques :

- **Julien Courtial**, pâtissier-chocolatier à Toumon-sur-Rhône,
- **Domaine des Collines** : Vins IGP Collines Rhodaniennes



**[Sam 21:30] Animation : 4TrakZ** 

Vivez un moment Alchimique avec **4TrakZ** !

Discographie :

- **Very Short Stories** LP0 Apr 08, 2013
- **Le Frère Lumière (Episode 1)** EP1 Aug 13, 2009




<http://4trakz.com/>




**[Sam 23:15] Fan-Film : Le Retour des Molten** 

Projection du fan-film de Benjamin Yoris (OneVision) : **Le Retour des Molten**.


Synopsis : Une mission de la Fédération des Planètes Unifiées étudiant la planète 3-7-0 dans le système Mikasa ne donne plus de signe de vie. L'équipage de l'Enterprise est chargé de découvrir ce qui s'est passé...



**[Dim 9:30] Le chipset graphique de l'Amiga** 

Le chipset graphique de l'Amiga.

Orateur : Victorien Ferry (Krabob)



2015 [revol...] [icons] [[nvda...] [[slee... [Images] Table ... [revol... [revol... [qemu... [revol... RMLL2... [revol... stands... slides... slides...



# References

- <http://weboob.org/>
- <http://git.symlink.me/?p=weboob/devel.git>
- <http://people.symlink.me/~rom1/blog/weboob/>



# Conclusion

- There are other ways to browse the web
- WebOOB puts it in a (nut)shell.
- Scraping can be fragile (depends on HTML)
- But sometimes it's the only solution
- And it saves a lot of time!
  
- Questions?

