# Weboob, the web suppository

Romain, Christophe

romain@weboob.org

RMLL, July 2010

# Overview

- Weboob shouldn’t exist
- Websites should export their data through (REST) webservices
- Benefits: access data without presentation, write complex treatments
- But they don’t.
- Problems:
  - Data can’t be fetched
  - Presentation can’t be modified
  - We are slave of the navigation

# Capabilities

Websites provide common patterns
We can extract generic interfaces: *capabilities*
Examples:

- video (youtube, dailymotion, youporn, break.com, etc.)
- bank (Crédit Agricole, BNP, Boursorama, etc.)
- weather (Yahoo! Weather, etc.)
- transports departues/arrivals (Transilien, RATP, voyages-sncf.com, etc.)
- e-commerce (ebay, priceminister, leboncoin, etc.)
- ...

# Backend

- Each website has it backend
- Each backend implements one or many capabilities
- Weboob provides tools to write backends

# Objects

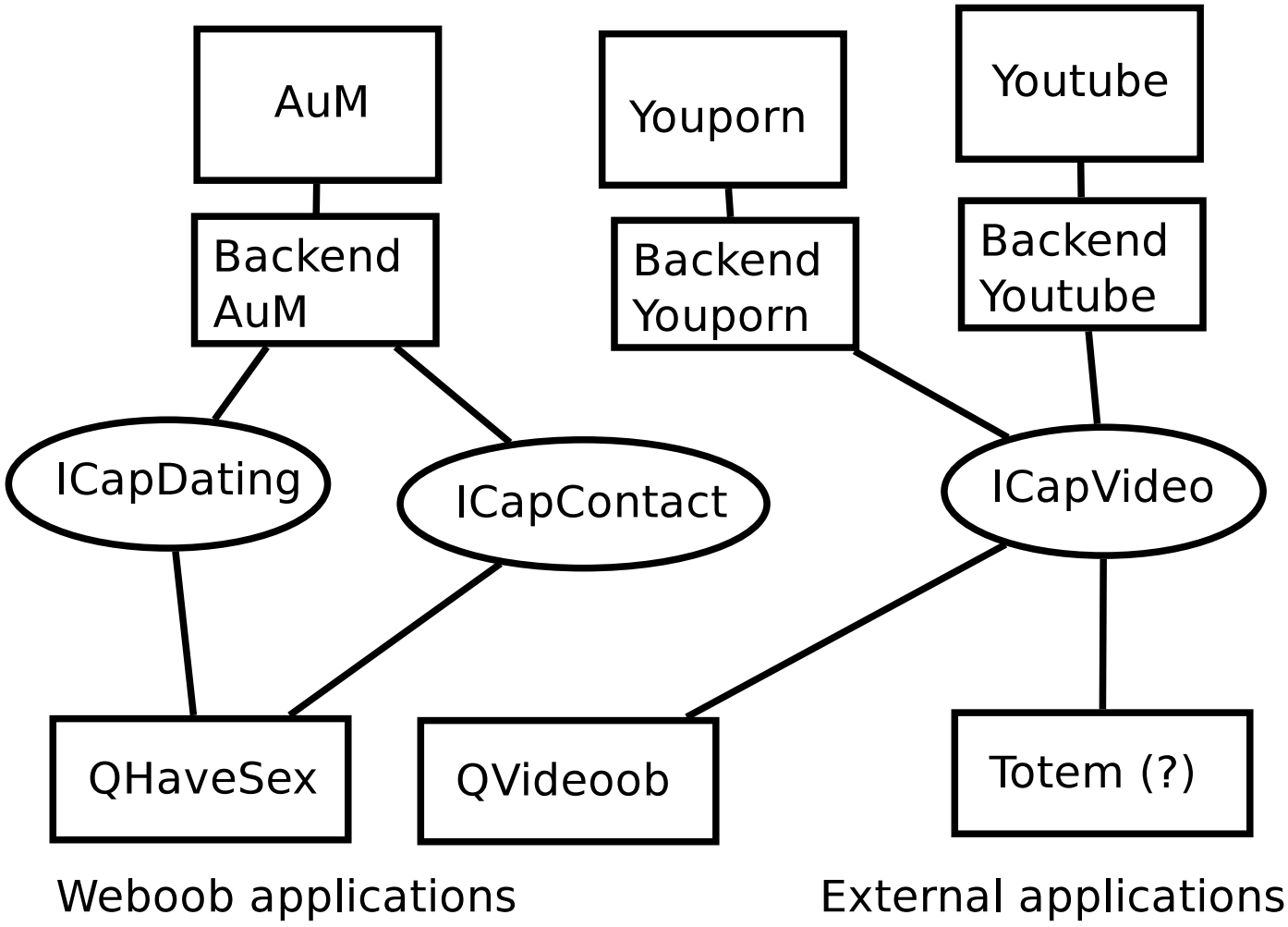
Example:

```
from weboob import Weboob
from weboob.capabilities.video import ICapVideo
weboob = Weboob()
weboob.load_modules(ICapVideo)
for backend, video in weboob.do('iter_search_results', pattern='coluche'):
    print video

id: qDENiP6lnWM@youtube
title: Italian Lessons
formatted_duration: 0:00:53
url: http://www.youtube.com/xxx/yyy.flv
author: Roberto546
thumbnail_url: http://i.ytimg.com/vi/qDENiP6lnWM/2.jpg
page_url: http://www.youtube.com/watch?v=qDENiP6lnWM
```

# Application

- Can be a daemon, console, GUI or web server
- Doesn’t know about backends, only capabilities
- Concentrate on display

# Diagram


AuM
Youporn
Youtube
Backend AuM
Backend Youporn
Backend Youtube
ICapDating
ICapContact
ICapVideo
QHaveSex
QVideoob
Totem (?)
Weboob applications
External applications

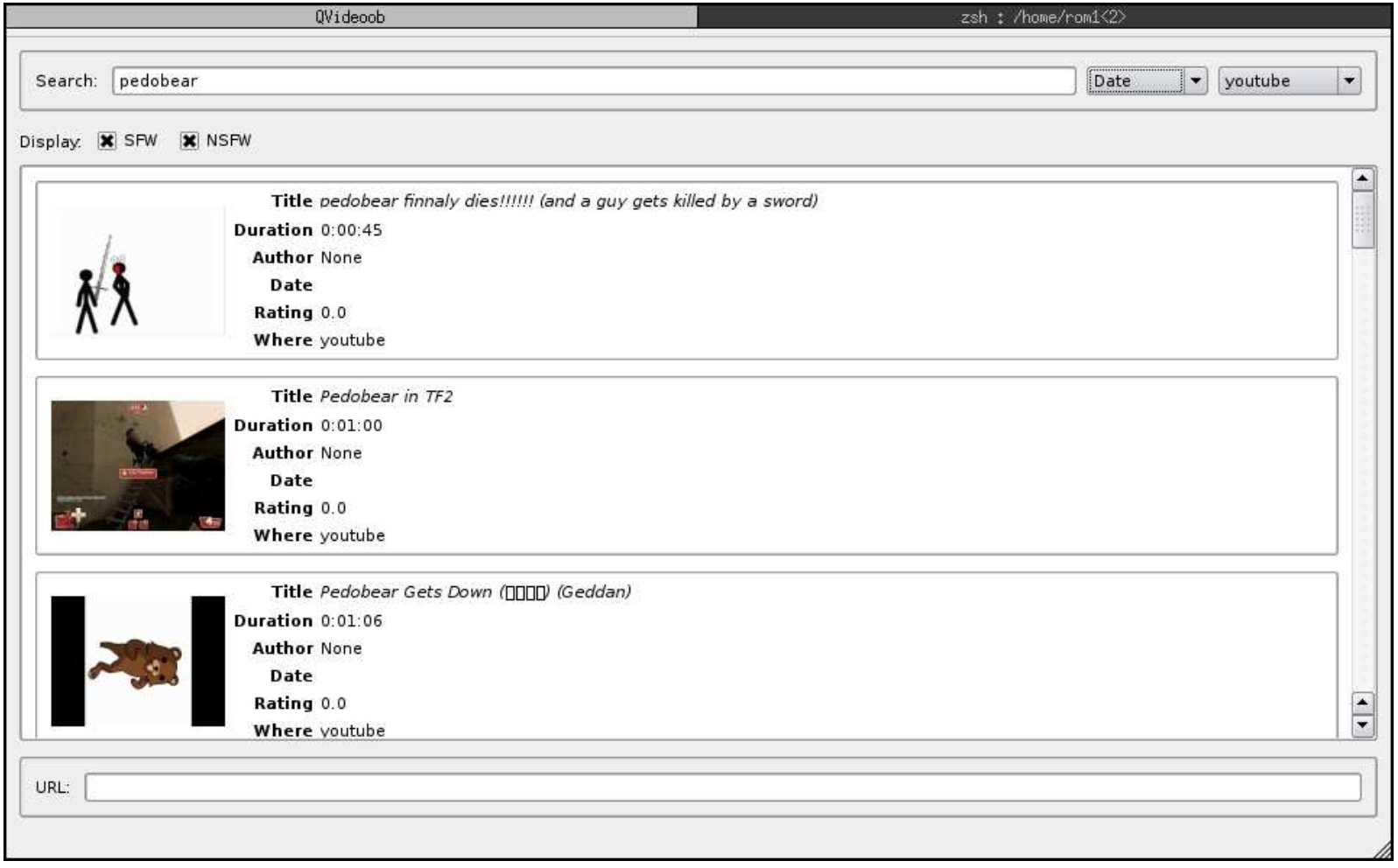
# ICapVideo

Example: video providers
Capability: ICapVideo
Many backends: youtube, youporn, INA, etc.
Same data: videos (title, description, duration, rating, etc.)
Same actions: search, get infos

# Videoob

- Search a video on many providers in parallel
  - $ videoob search candy
- Get information about a video
  - $ videoob info JPONTneuaF4@youtube
  - $ videoob info JPONTneuaF4@youtube --select url
  - $ videoob info JPONTneuaF4@youtube --format webkit

# QVideoob


QVideoob
zsh : /home/rom1<2>
Search: pedobear
Date
youtube
Display: SFW NSFW
Title pedobear finnaly dies!!!!!! (and a guy gets killed by a sword)
Duration 0:00:45
Author None
Date
Rating 0.0
Where youtube
Title Pedobear in TF2
Duration 0:01:00
Author None
Date
Rating 0.0
Where youtube
Title Pedobear Gets Down (□□□□) (Geddan)
Duration 0:01:06
Author None
Date
Rating 0.0
Where youtube
URL:

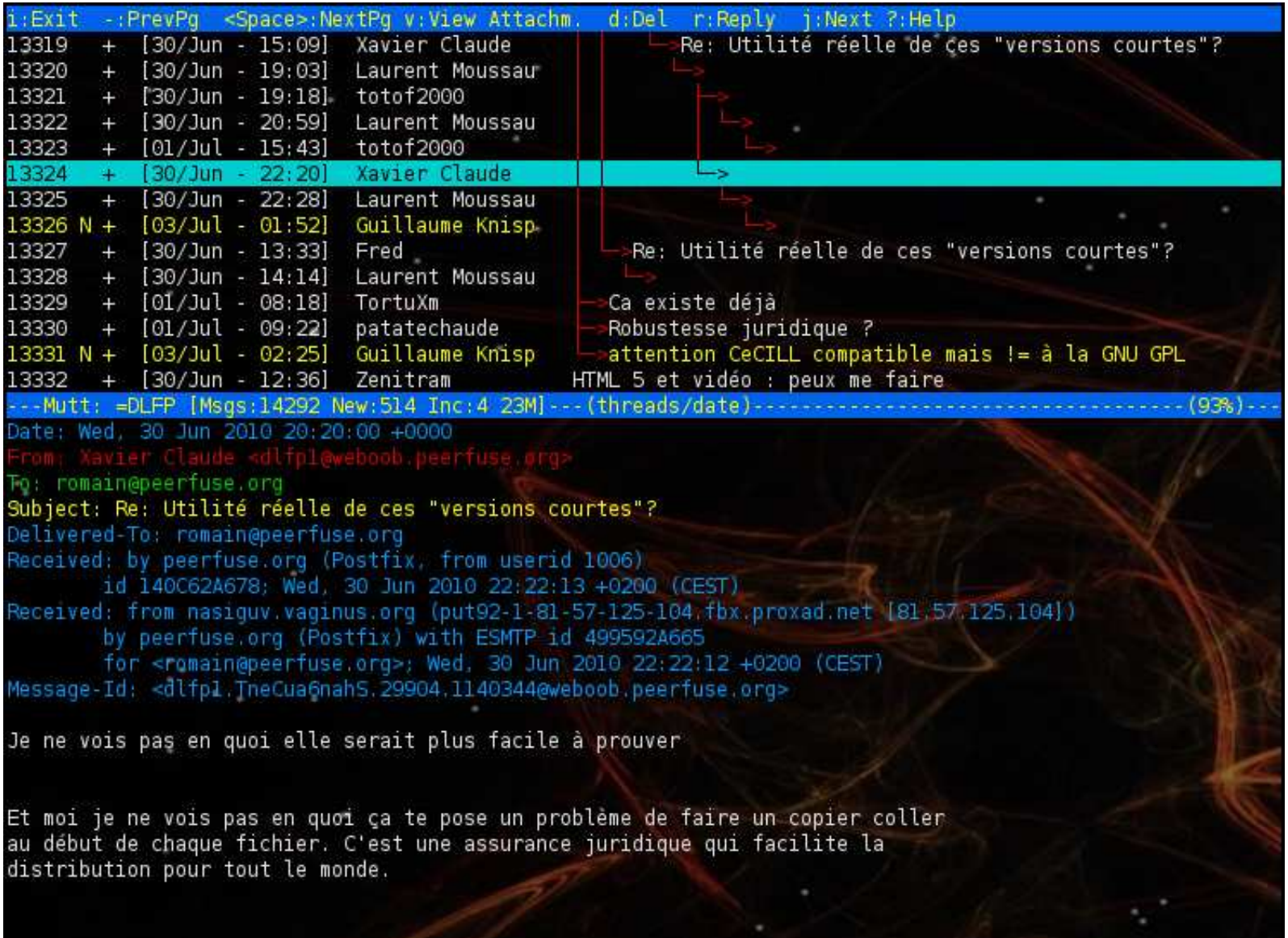
# Weboorrents

```
$ weboorrents search debian
seeders: 15
description: Debian Lenny netinstall iso
leechers: 9
date: None
size: 321545830.4
id: 1578687@mytracker
name: debian-40r3-amd64-netinst.iso

[...]

$ weboorrents getfile 1578687@mytracker ~/Watch/1/debian.torrent
$
```

# Boobank

```
$ boobank list -f table
+--------------------+----------+--------------------------+---------+
|       label        | balance  |            id            |  coming |
+--------------------+----------+--------------------------+---------+
| Compte de chèques | 2878.12  | 01255000004XXXXX@bnporc | -333.31 |
| Livret Jeune       | 674.17   | 01255000740XXXXX@bnporc | 0.0     |
+--------------------+----------+--------------------------+---------+
```

# Monboob


i:Exit -:PrevPg <Space>:NextPg v:View Attachm. d:Del r:Reply j:Next ?:Help
13319 + [30/Jun - 15:09] Xavier Claude Re: Utilité réelle de ces "versions courtes"?
13320 + [30/Jun - 19:03] Laurent Moussau
13321 + [30/Jun - 19:18] totof2000
13322 + [30/Jun - 20:59] Laurent Moussau
13323 + [01/Jul - 15:43] totof2000
13324 + [30/Jun - 22:20] Xavier Claude
13325 + [30/Jun - 22:28] Laurent Moussau
13326 N + [03/Jul - 01:52] Guillaume Knisp
13327 + [30/Jun - 13:33] Fred Re: Utilité réelle de ces "versions courtes"?
13328 + [30/Jun - 14:14] Laurent Moussau
13329 + [01/Jul - 08:18] TortuXm Ca existe déjà
13330 + [01/Jul - 09:22] patatechaude Robustesse juridique ?
13331 N + [03/Jul - 02:25] Guillaume Knisp attention CeCILL compatible mais != à la GNU GPL
13332 + [30/Jun - 12:36] Zenitram HTML 5 et vidéo : peux me faire
---Mutt: =DLFP [Msgs:14292 New:514 Inc:4 23M]---(threads/date)---(93%)---
Date: Wed, 30 Jun 2010 20:20:00 +0000
From: Xavier Claude <dlfp1@weboob.peerfuse.org>
To: romain@peerfuse.org
Subject: Re: Utilité réelle de ces "versions courtes"?
Delivered-To: romain@peerfuse.org
Received: by peerfuse.org (Postfix, from userid 1006)
id 140C62A678; Wed, 30 Jun 2010 22:22:13 +0200 (CEST)
Received: from nasiguv.vaginus.org (put92-1-81-57-125-104.fbx.proxad.net [81.57.125.104])
by peerfuse.org (Postfix) with ESMTP id 499592A665
for <romain@peerfuse.org>; Wed, 30 Jun 2010 22:22:12 +0200 (CEST)
Message-Id: <dlfp1.JneCuaGnahS.29904.1140344@weboob.peerfuse.org>
Je ne vois pas en quoi elle serait plus facile à prouver
Et moi je ne vois pas en quoi ça te pose un problème de faire un copier coller
au début de chaque fichier. C'est une assurance juridique qui facilite la
distribution pour tout le monde.

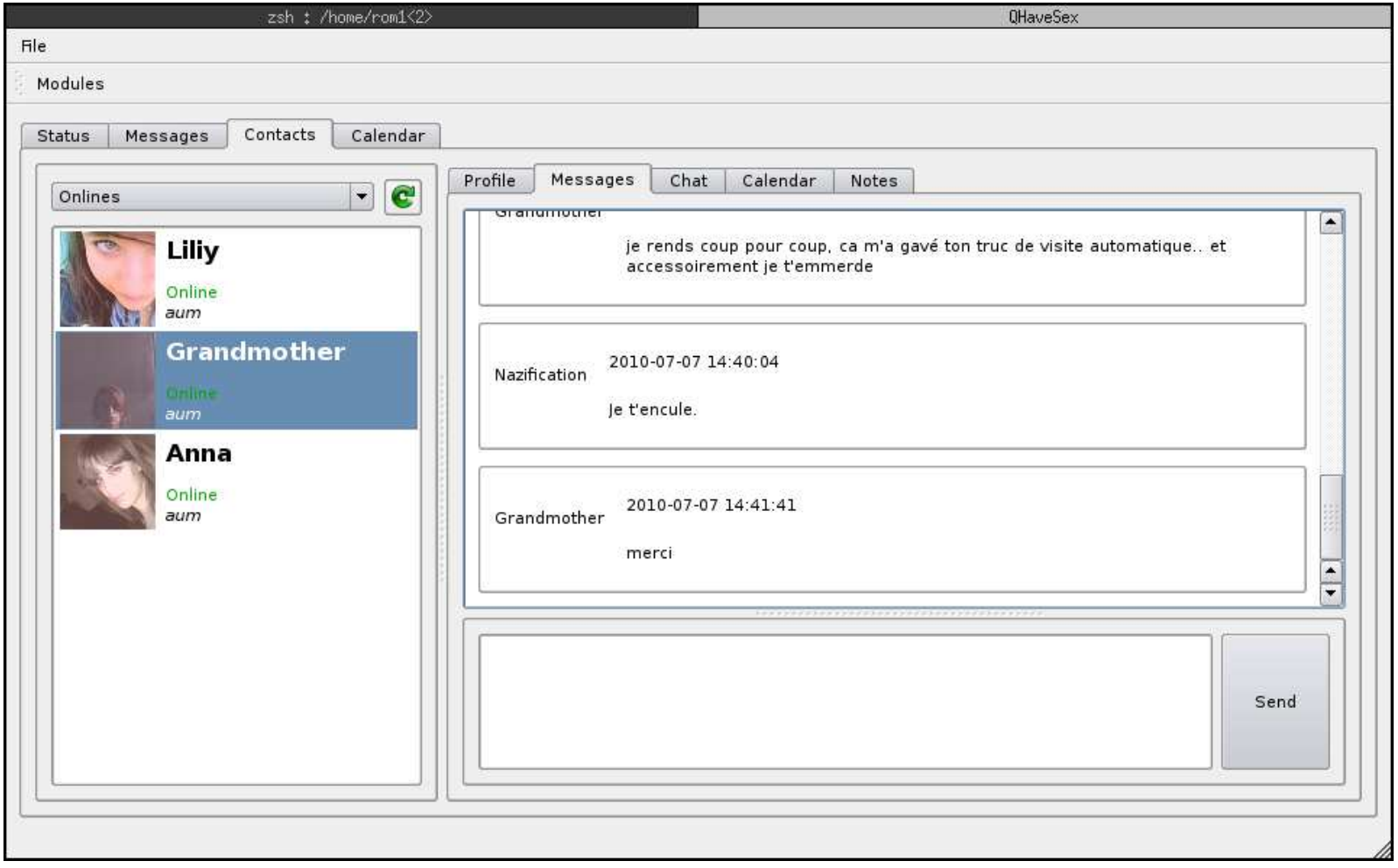
# MassTransit

- N900 application
- Use the hildon framework
- Departures from a station

# QHaveSex

- Contacts management
- Exchange messages
- Automation of the drag process
- Calendar to plan dates
- Places management

# QHaveSex


zsh : /home/rom1<2>
QHaveSex
File
Modules
Status
Messages
Contacts
Calendar
Onlines
Liliy
Online
aum
Grandmother
Online
aum
Anna
Online
aum
Profile
Messages
Chat
Calendar
Notes
je rends coup pour coup, ca m'a gavé ton truc de visite automatique.. et accessoirement je t'emmerde
Nazification
2010-07-07 14:40:04
Je t'encule.
Grandmother
2010-07-07 14:41:41
merci
Send

- Backends implement capabilities
- Each backend can have a browser, based on mechanize
- The browser parses the HTML and returns capabilities objects
- Applications receive these objects and treat them like other backends objects

Browsers are independent of weboob.

Howto develop a backend?

- Choose an existing capability to implement
- or create a new capability
- Examine the website with firebug to understand how it works
- Write the new backend, the new browser, and return filled objects
- Now, you can use the existing applications, which call the new backend

- Simple browsers (urlopen + parse HTML string with regexps)
- Simple browsers (urlopen + parse HTML string to DOM)
- Complex browsers which know website pages organization

# Future

- Unit tests to detect broken backends
- Provide tools to help writing backends
- Rewrite the core library in C

# Application ideas

- E-commerce best price search
- Housing search (with some AI)
- Search for music concerts on artists websites

# Questions?