Robust adaptive discourse parsing for e-learning fora

Nadine Lucas & Emmanuel Giguet

Cnrs  Caen University France

http://www.info.unicaen.fr/~nadine
Outline

• Context
• “Agora” forum parsing principles
• Results
• Example: parsing on the fly
• Conclusion
Main objectives

• Follow-up of students’ fora (on-line discussions)
  – Monitoring the students’ participation
  – Detecting the cold start problem
  – Detecting building up of momentum in collective discussion

• Reflection on past experience
  – Tutor’s intervention

• Give access to content (text itself)
What is the problem?

• Large amount of textual data
  – Scrolling and reading takes time

• Yet, sentence parsing is not efficient
Words in sentences?

Code: . other-script.sh
OR

Code: . ./other-script.sh
OR

Code: ./other-script.sh
I get other-script.sh: No such file or directory. I already chmod'ed, so other-script is readable and executable. Any ideas?

#8 07/02/2008 13:41 Mike
Re - Re: Running a script inside another
max wrote: Code: ./other-script.sh

This one should work, Max. Maybe for testing purposes, put the full path to the script in there?

#9 07/02/2008 14:07 max [is back]
Re -
found the problem...i had cd'ed into another directory and then tried to run the script with only a relative pathname.

#10 07/02/2008 16:29 immunity
Project 1 Question on command line restriction
So I remembered I was going to ask you this literally 1 minute after you left the room:

So lets say we're using the default command line

Code: ./fetch-rss.sh -i http://twitter.com/statuses/user_timeline/12104172.rss -o rss-output

Are we supposed to have this command restricted to where -i and -o HAVE to be in those spots, or are we supposed to give them the ability to switch it up, having the directory destination specification first, then the URL second.

Or is the choice ours?

Thanks,
Scale related to expectations

• 15 fora going on at the same time on a platform
  – 53 threads in a forum and 166 posts
• Have a look on how the forum is faring
  – Assess collaboration
• Discourse parsing?
  – Meaning units?
Calico

- Calico (French Ministry of Education)  
  - 2005-2008
- Practitioners and researchers  
  - 10 teams
- Exchange platform  
  - https://wims.crashdump.net/www/calico/
- Agora forum parser is one among many tools
Monitoring tools
E-learning

- Students’ on-line discussions (BBs, fora)
  - Distance learning
  - Presence learning
  - Mixed
- French, English, Spanish
<table>
<thead>
<tr>
<th>THEME</th>
<th>RHEME</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boujour Malvina et Gwenaelle :D</strong> Je propose qu'on discute un peu sur les deux scenarios qui nous étaient venus à l'esprit. [...] je comprends</td>
<td><strong>Bonjour, [...] Séphane Budoux</strong></td>
</tr>
<tr>
<td><strong>C'est juste 2 exemples d'apprentissages interactifs que j'avais crée pour mon propre usage. [...] qui fait ?</strong></td>
<td><strong>Si donc vous modifiez les pages en graphique, [...] ---------------------------------------- &lt;a href='http://azurereole.free.fr' target='_blank'&gt;<a href="http://azurereole.free.fr">http://azurereole.free.fr</a>&lt;/a&gt;</strong></td>
</tr>
<tr>
<td><strong>Apparemment, [...] j'ai en tête le fichier Word qui est peut être plus facile à retrouver sur son disque que cette post.</strong></td>
<td><strong>En reliant le fichier de vendredi, [...] Ce code &quot;parait&quot; brouille rapidement la lecture et génère considérablement les interventions manuelles dans le fichier.</strong></td>
</tr>
<tr>
<td><strong>En reliant le fichier de vendredi, [...] Ce code &quot;parait&quot; brouille rapidement la lecture et génère considérablement les interventions manuelles dans le fichier.</strong></td>
<td><strong>En reliant le fichier de vendredi, [...] Ce code &quot;parait&quot; brouille rapidement la lecture et génère considérablement les interventions manuelles dans le fichier.</strong></td>
</tr>
<tr>
<td><strong>Fait plan de mise en page du cahier des charges sur azurereole afin de permettre aux commentaires de mieux s'espacoir. [...] je le ferai le même chose sur le page scenario quand on y travaillera.</strong></td>
<td><strong>Finalement, [...] Question sur laquelle on parle de mettre en place le relevé de la phase en présentiel plan de travail préparé, [...] mais ça tombe sous le sens.</strong></td>
</tr>
<tr>
<td><strong>Finalement, [...] Question sur laquelle on parle de mettre en place le relevé de la phase en présentiel plan de travail préparé, [...] mais ça tombe sous le sens.</strong></td>
<td><strong>La date butoir c'est le 01/11, [...] a plus</strong></td>
</tr>
<tr>
<td><strong>Salut :D je résume donc : je vais faire une copie du cdc en étal et avez commentaires sur forum dép. [...] Voir pièce jointe</strong></td>
<td><strong>Bonjour, je viens de lire l'évolution du cahier des charges, [...] +</strong></td>
</tr>
<tr>
<td><strong>Bonjour, je viens de lire l'évolution du cahier des charges, [...] +</strong></td>
<td><strong>A force de remplanter les textes petits bouts par petits bouts, [...] même si il ne s'agit que de remplacer un fichier sous forme clairement visible.</strong></td>
</tr>
<tr>
<td><strong>A force de remplanter les textes petits bouts par petits bouts, [...] même si il ne s'agit que de remplacer un fichier sous forme clairement visible.</strong></td>
<td><strong>Bonjour, éh bien tant pis pour le délai. [...] Des exemples de ce que j'appelle fichiers illisibles sont disponibles sur le site pour tous les fichiers retrouvables sous le dossier de graphique.</strong></td>
</tr>
<tr>
<td><strong>La page --&gt; [...] &lt;a href='http://azurereole.free.fr' target='_blank'&gt;<a href="http://azurereole.free.fr">http://azurereole.free.fr</a>&lt;/a&gt;</strong></td>
<td><strong>&lt;br&gt; [...] je valide à priori et par avance vos modifications du cahier des charges</strong></td>
</tr>
<tr>
<td><strong>Je sais que toi, [...] ------------------------------ La précision de ce que j'entends par détails ou par test se trouve par contre dans les messages jusque au dessus</strong></td>
<td><strong>alors ok, gardons les tests avant le 15 mais pourquoi se berner sur des modifications qui pourraient être judicieuses après cette date?</strong></td>
</tr>
<tr>
<td><strong>alors ok, gardons les tests avant le 15 mais pourquoi se berner sur des modifications qui pourraient être judicieuses après cette date?</strong></td>
<td><strong>Rien ne nous empêche de réviser des éléments après cette date butoir, [...] +</strong></td>
</tr>
<tr>
<td><strong>Rien ne nous empêche de réviser des éléments après cette date butoir, [...] +</strong></td>
<td><strong>Je sais que toi, [...] ------------------------------ La précision de ce que j'entends par détails ou par test se trouve par contre dans les messages jusque au dessus</strong></td>
</tr>
<tr>
<td><strong>Je sais que toi, [...] ------------------------------ La précision de ce que j'entends par détails ou par test se trouve par contre dans les messages jusque au dessus</strong></td>
<td><strong>c'est juste sur le motif qu'il n'était pas compris, que tu as prête au dossier ( beaucoup plus clair pour moi. [ ... ] j'en suis désolé que mon respect envers toi t'importunne autant</strong></td>
</tr>
<tr>
<td><strong>c'est juste sur le motif qu'il n'était pas compris, que tu as prête au dossier ( beaucoup plus clair pour moi. [ ... ] j'en suis désolé que mon respect envers toi t'importunne autant</strong></td>
<td><strong>la page --&gt; [...] je précise que je valide à priori et par avance vos modifications du cahier des charges.</strong></td>
</tr>
<tr>
<td><strong>la page --&gt; [...] je précise que je valide à priori et par avance vos modifications du cahier des charges.</strong></td>
<td><strong>Je ne comprends pas ta réaction. [...] + J'ai effectivement reçu ce type de réactions et les ai intégrés comme prévu au travail en cours. - J'apprecie d'ailleurs que tu aies tenu compte de mes suggestions.</strong></td>
</tr>
<tr>
<td><strong>Je ne comprends pas ta réaction. [...] + J'ai effectivement reçu ce type de réactions et les ai intégrés comme prévu au travail en cours. - J'apprecie d'ailleurs que tu aies tenu compte de mes suggestions.</strong></td>
<td><strong>Je ne comprends pas ta réaction. [...] + J'ai effectivement reçu ce type de réactions et les ai intégrés comme prévu au travail en cours. - J'apprecie d'ailleurs que tu aies tenu compte de mes suggestions.</strong></td>
</tr>
<tr>
<td><strong>Je ne comprends pas ta réaction. [...] + J'ai effectivement reçu ce type de réactions et les ai intégrés comme prévu au travail en cours. - J'apprecie d'ailleurs que tu aies tenu compte de mes suggestions.</strong></td>
<td><strong>Et bien si ma réaction est due à la fatigue, [...] Alors je me borne à rappeler l'essentiel : Je valide à priori et par avance vos modifications du cahier des charges et plus je finis.</strong></td>
</tr>
<tr>
<td><strong>Et bien si ma réaction est due à la fatigue, [...] Alors je me borne à rappeler l'essentiel : Je valide à priori et par avance vos modifications du cahier des charges et plus je finis.</strong></td>
<td><strong>Mes excuses pour le code apparemment non compréhensible que j'ai fourni pour la réforme du cahier des charges de mardi dernier, [...] Ceci afin de valider l'étape.</strong></td>
</tr>
</tbody>
</table>
Agora

Input whole forum file html

Conversion to XML

Segmentation

Chrono order

Parsing

Visualisation

Output coloured hierarchy
Agora parsing principles

• On line discussion
  – Collective discourse

• Time line
  – Rhythm

• Projected interpretation grid
  – Expository discourse + communication

• Difference principles
Rythm

• Start versus discussion proper
  – Coordination and subordination relations
  – By default three levels
### 3 levels

#### G.1

**G.1** So... P.S.-I'm sure this was just a typo, but alamode is misspelled in the first ssh command.

Yes, it was a typo. But suffice it to say that if you replace start and end with the opening and closing item tags then it generally works on the xml files.

#### G.2

**G.2** There isn't sample code on Ore about getopts, so here is a program that accepts -a arg or -b arg, but not both: You can use it to start with and then maintain it, or you can use getcwd(2) instead.

KDevelop doesn't like to install when you're using Virtual PC. I've wondered where this started and I think it goes back to the time I went to the circus and a clown killed my dad.

#### G.3

**G.3** Wow. That's an impressive splash! You open a file, fork, and in the child dup the open file's file descriptor onto standard in, the call exec. There aren't any pipes involved.

#### G.3.1

**G.3.1** In running through the tests listed in the tests post, I ran this test (from Sara): Do we need to capture this situation properly, or can I call my bug a feature?

#### G.3.2

**G.3.2** Mine is super fancy... I most certainly will be fishing from now on Brandon Cruz Ardiente Like R-D-N-T

#### G.3.3

**G.3.3** Or, you could scope lock on the cout statement... Hopefully that helps a little bit. -Sara McF
Find the odd element in a series

- Whole forum (at time T)
  - Background pattern
    - Standard message length and structure
    - Standard exchange structure
  - Salient features
    - Odd post(s) in a series
    - Border
Relative saliency

- Detection of similarities or differences
  - Along time
    - related features, same patterns --> coordinate
  - According to distributional saliency
    - new patterns --> subordinate or superordinate
    - hierarchy in inverse frequency
<table>
<thead>
<tr>
<th>Author</th>
<th>Message</th>
</tr>
</thead>
</table>
| smcفارلا    | I was wondering whether our code should allow for requests to have the same name, or whether this shouldn't be allowed. For example, if the following were entered: Code:  
```
timer-server>> foo 1 2  
timer-server>> foo 1 2  
timer-server>> foo 3 2  
```  
Is the program expected to show the statuses of all of these when "stat" is entered? Or should we not allow the second and third requests because that name is already present?  
-Sara McF                                                              |

<table>
<thead>
<tr>
<th>Author</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Admin</td>
<td>No, don't allow the second and third requests.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Admin</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Admin</td>
<td></td>
</tr>
</tbody>
</table>
Relative difference

• No exhaustive description
• Just check differences
  – Message groups homogeneity
    • Message size
    • Message structure
  – Distribution of rare contrastive salient features
    • HTML labels
    • Smilies, punctuation
Technical side

• XMLForum exchange format
• Segmentation
• Chronological ordering
• Parsing
• Visualisation
# 1 01/02/2008 12:32 max (e back)
Bash best practices
I know it's important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? My guess is yes, but it seems to make my code a little bloated sometimes.

For instance, if you have a function that takes a directory as an argument, not only would you have to check that the number of arguments is correct, but also that the first argument really is a directory.

# 1 01/02/2008 13:10 Dan Lecocq
Shell regexps
My shell is being really finnicky (seemingly): it makes me use * instead of +. Is this a standard of some sort? I'm used to * being 0-or-more and + being 1-or-more, and it's even been weird about me using ? (although, * isn't greedy as I understand it).

At any rate, I'm used to regexps in Ruby (I think they're just like in perl), but I'm just wondering if my shell is weird (I'm running bash in OS X), or if I'm not doing something right. So far it hasn't been a problem per se, just disconcerting.

# 1 01/02/2008 13:44 Stuart Fehr
Re -
Your shell is not using true regular expressions. Rather, it is using something called glob. See http://en.wikipedia.org/wiki/Glob_%28programming%29 for a terse description of glob. The syntax of what is interpreted with globbing is much more limited than full regular expressions, but there are many good resources online for determining what is or is not allowed in your glob expression.

# 2 01/02/2008 14:10 Mike Colagrosso
Re -
Stuart is exactly right.
I forget to mention it in class yesterday, but I put four books in the Makebelab that will help you with it.

Learning the bash shell
Unix in a Nutshell
Debating Foods with RSS and Atom
There are also a book in there on sed and awk, plus a few more that you might like.

# 2 01/02/2008 14:38 Mike Colagrosso
Re -
I'm glad that you asked this, Max, because it brings up some conflicting ideas. By the title of your post I thought you might be checking the arguments of a script. If so, your way is fine.

One reason for checking the arguments of scripts and functions is to program defensively, and the other is to give a way to check if you should do an awk of maybe it works. "What is the point of you finding yourself calling 'ls -a' twice possibly for lots of functions, I would suggest you check if you're interested in one more bit about my style, I tend to write few functions in shell scripts, and the

# 6 01/02/2008 16:14 Mike Colagrosso
Do you follow me?
Here are some Twitter accounts to subscribe to in order to test your Project 1.

These are updated frequently, so they will test your code's ability to get new posts:
John Gruber
Merlin Mann
You might like an account that brings the funny:
Vader
This is the account that I post to, I update it when something that the MobiSys audience would be interested to. I hooked these updates into the conference website.
MobiSys 08
There's many more worth suggesting, including at least one account from one of our classmates.
Wrappers and snippets

#1 01/02/2008 12:32 max [is back]
Bash best practices
I know it's important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? My guess is yes, but it seems to make my code a little bloated sometimes.

For instance, if you have a function that takes a directory as an argument, not only would you have to check that the number of arguments is correct, but also that the first argument really is a directory.

#2 01/02/2008 13:10 Dar
Shell regexps
My shell is being really finnicky (seemingly) makes me use * instead of +. Is this a standard of some sort? I'm used to * being 0-or-more and + being 1-or-more, and it's even been weird about me using ? (although, * isn't greedy as I understand it).

At any rate, I'm used to regexps in Ruby (I think they're just like in perl), but I'm just wondering if my shell is weird (I'm running bash in OS X), or if I'm not doing something right. So far it hasn't been a problem per se, but just disconcerting.

#3 01/02/2008 13:44 Stuart Fehr

Your shell is not using true regular expressions. Rather, it is using something called glob. See http://en.wikipedia.org/wiki/Glob_%28programming%29 for a terse description of glob. The syntax of what is interpreted with globbing is much more limited than full regular expressions, but there are many good resources online for determining what is or is not allowed in your glob expression.

These are updated frequently, so they will test your code's ability to get new posts:

- John Gruber
- Merlin Mann

You might like an account that brings the funny:

- Vader

This is the account that I post to. I update it when something that the MobiSys audience would be interested to. I hooked these updates into the conference website.

MobiSys 08

There's many more worth suggesting, including at least one account from one of our classmates.
I know it's important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? [...] So far it hasn't been a problem per se, but just disconcerting.
Visualisation

• Show compact view
  - Tuning *versus* Discussion proper
  - Discussion divided in “moments”
    • Not topics

• Zooming in
  - Moments sub divided in rounds

• All units expandable
  - Showing full content
CHARSET DETECTOR OUTPUT =

**G**
I know it's important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? [...] Maybe for testing purposes, put the full path to the script in there?

**G.1**
found the problem...i had cd'ed into another directory and then tried to run the script with only a relative pathname. [...] My only thought was set the PATH variable to be the main directory containing all my code subdirectories and then somehow make the user run it from there.

**G.2**
Matt, I don't think the ending slash is an issue with the -o option to fetch-rss. If you call Code: mkdir some-directory-name/ it will nuke the final slash and create a directory called some-directory-name.
Results

• Show only main hierarchy
  – Provide a kind of signature for fora

• Compare fora at a glance
  – on the same period or same task
  – for different classes or different groups
## OS Projects 07 vs 08

**BG**

| G | Code: cat index.xml | grep enclosure | sed 's/\*url="/\*"/\*$/1/' You should work on command-line arguments after you have the basic functionality of your fetch-rss.sh code working. Another thing that I found insightful were the handy links at the bottom of this page. If you click on the Valid HTML and Valid CSS links, those pages can give you details about what is wrong with your page. |
|---|---|
| G | [2] You can use it to start with and then maintain it, or you can use getcwd(2) instead. |

**G.1**

So... [2] P.S.-I'm sure this was just a typo, but alamode is misspelled in the first ssh command.

**G.2**

There isn’t sample code on Ore about getoptps, so here is a program that accepts -a arg or -b arg, but not both: [3] You can use it to start with and then maintain it, or you can use getcwd(2) instead.

**G.3**

Wow. That’s an impressive splash! You open a file, fork, and in the child dup the open file’s file descriptor onto standard in, the call exec. There aren’t any pipes involved.

**BG**

| G | I know it’s important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? My guess is yes, but it seems to make my code a little bloated sometimes. [2] Jason Hopkins wrote: Code: VAR="' awk 'string {print} filename | tr 1 2" This code works for me, Jason, although it doesn’t print anything. What is it supposed to do? |
|---|---|
| G | Dan and I worked on some to convert the dates in the RSS feeds to something that is easy to sort: Code: $ date -d "Fri, 11 Jan 2008 19:49:49 +0000" "%y-%m-%d-%H-%M-%S" 08-01-11-12-49-49 and here’s a version that works on Mac OS X: Code: $ date -j -f "%a, %d %b %Cyy %H:%M:%S %z" "Fri, 11 Jan 2008 19:49:49 +0000" "+%y-%m-%d-%H-%M-%S" 08-01-11-12-49-49 so I’m trying to delete an entire directory with contents in it, yet rd doesn’t work in alaMode. On the contrary, everything inside of the $( ) is interpreted normally just as if it were its own command. (Also, it is possible to jump from the beginning to the end or vice-versa of one of these blocks in vim using % while that’s not possible with the backquotes.) On a side note, the section that details this also suggests that if you need to do Code: $ blah $(cat file) you instead use Code: $ blah $(< file) as it is more efficient. This is because bash handles opening and reading the file it instead of creating a new process to do this. (Don’t worry you’ll get to creating and destroying processes and switching between them later in the semester.) |
| G | their know thyme lichen thee presence. |

**G.2**

Thanks for sharing this Dan. I need to get a few things fixed in my code before I give it a go, but I’m glad I have this. [2] -m

**G.3**

We tried gdb, we tried make clean and make, and yet... It seems to be breaking with the bit in Player.php about: Code: gc->turnaround_times.push_back(turnaround); I’ve not seen anything so bizarre in recent memory. [2] I may be getting ridiculously incredibly lucky. Hope this helps.

**G.4**

I was hoping to hear some chatter on this project. Chris, I can look at this more closely when I return, but what I think is going on is that you need to make sure your Python files get packaged up and sent to the slave machines. [2] Edited by Mike C.: Please see my note below about getting the file from S3.
OS Concepts ≠ OS Projects 07

I'll use this topic to post announcements of interest to the class. You can always stay up to date by having the board notify you if I post something here. At the bottom of this page, there's a link that reads "Watch this topic for replies." Click it, and Ore will email you if I post something new. [1] For those who are interested, the MLRC lab is not on ADT, so if, since we're getting the Alamo login first, you run both passwd and smbpasswd on an Alamo machine, it will affect both Alamo and MLRC. "smbpasswd" is the critical one - passwd will change your UNIX password (i.e. Alamo), "smbpasswd" on the other hand, modifies your Samba password - Samba is the protocol used for Windows sharing on UNIX. Therefore, smbpasswd will modify your password for Windows access - and, logically, the MLRC lab. (Verification from ADS staff - all blame lies on him if it doesn't work!!!)  ________________ Ryan Carpenter ADS Student Consultant/Mines Help Center rcarpent@mines.edu 719/930-3412

G.1 Taken/interpreted from Wepedia (http://www.webopedia.com/TERM/b/big_endian.html)... [...] Very nice.

G.2 What is the purpose of System Calls and how do system calls relate to the operating system and to the concept of dual-mode (kernel mode and user mode) operations? System calls allow user-level processes to request services from the operating system that the process itself is not allowed to do. These system calls are classified into broad six categories - Filesystem, Process, Scheduling, Interprocess communication, Networking, and Miscellaneous. [...] Code: du -s disk usage by you (assuming you are in your home directory) Host name: Code: uname -n OS: Code: uname -o

G.3 Let's try a different way to explain this one. [...] Later, Tom Brokaw shows up and enjoys two wonderful meals.

G.4 Brandon Ardiente wrote: After the test on Tuesday and tonight's programming project, I'm definitely feeling devious. [...] I'd like to say no, but I don't know the reasons why not.

G.5 Critique of paging: [...] Yeah, I also added the export line to my .bashrc (and ran it in the terminal) before I tried make and I got the same results as Brandon. ________________ -Sara McF

G.6 Teresa Davies wrote: I've found that when I ssh into alamode, I have to set LD_LIBRARY_PATH every time. [...] You'll use the readline and Boost threading libraries in this project.

G Code: cat index.xml | grep enclosure | sed 's/\^.*url="([^\^]*).*/\1/' You should work on command-line arguments after you have the basic functionality of your fetch-rss.sh code working. Another thing that I found insightful were the handy links at the bottom of this page. If you click on the Valid HTML and Valid CSS links, those pages can give you details about what is wrong with your page. ________________ -Sara McF

G.1 So... [...] P.S. I'm sure this was just a typo, but alamode is misspelled in the first ssh command.

G.2 There isn't sample code on Ore about getopt, so here is a program that accepts -a arg or -b arg, but not both: [...] You can use it to start with and then maintain it, or you can use getcwd(2) instead.

G.3 Wow. That's an impressive splash! You open a file, fork, and in the child dup the open file's file descriptor onto standard in, the call exec. There aren't any pipes involved.
### Zooming on OS Projects 07

<table>
<thead>
<tr>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G.1</strong></td>
</tr>
<tr>
<td><strong>G.1</strong> So... P.S. - I'm sure this was just a typo, but alamode is mispelled in the first ssh command.</td>
</tr>
<tr>
<td>Yes, it was a typo. But suffice it to say that if you replace start and end with the opening and closing item tags then it generally works on the xml files.</td>
</tr>
<tr>
<td><strong>G.2</strong></td>
</tr>
<tr>
<td><strong>G.2</strong> There isn't sample code on Ore about getopt, so here is a program that accepts -a arg or -b arg, but not both: You can use it to start with and then maintain it, or you can use getcwd instead.</td>
</tr>
<tr>
<td>KDEvelop doesn't like to install when you're using Virtual PC. I've wondered where this started and I think it goes back to the time I went to the circus and a clown killed my dad.</td>
</tr>
<tr>
<td><strong>G.3</strong></td>
</tr>
<tr>
<td><strong>G.3</strong> Wow. That's an impressive splash! You open a file, fork, and in the child dup the open file's file descriptor onto standard in, the call exec. There aren't any pipes involved.</td>
</tr>
<tr>
<td><strong>G.3.1</strong></td>
</tr>
<tr>
<td><strong>G.3.1</strong> In running through the tests listed in the tests post, I ran this test (from Sara): Do we need to capture this situation properly, or can I call my bug a feature?</td>
</tr>
<tr>
<td><strong>G.3.2</strong></td>
</tr>
<tr>
<td><strong>G.3.2</strong> Mine is super fancy... I most certainly will be fishing from now on Like R-D-N-T</td>
</tr>
<tr>
<td><strong>G.3.3</strong></td>
</tr>
<tr>
<td><strong>G.3.3</strong> Or, you could scope lock on the cout statement... Hopefully that helps a little bit. -Sara McF</td>
</tr>
</tbody>
</table>
Zooming on OS Projects 08

G.1 Dan and I worked on some to convert the dates in the RSS feeds to something that is easy to sort. Code: `date -d "Fri, 11 Jan 2008 19:49:49 +0000" "+%y-%m-%d-%H-%M-%S" 08-01-11-12-49-49 and here's a version that works on Mac OS X: Code: `date -j -f "%a, %d %b %C%y %H:%M:%S %z" "Fri, 11 Jan 2008 19:49:49 +0000" "+%y-%m-%d-%H-%M-%S" 08-01-11-12-49-49 so I'm trying to delete an entire directory with contents in it, yet rd doesn't work in alamode. On the contrary, everything inside of the $() is interpreted normally just as if it were its own command. (Also, it is possible to jump from the beginning to the end or vice-versa of one of these blocks in vim using % while that's not possible with the backquotes.) On a side note, the section that details this also suggests that if you need to do Code: `println $(cat file)` you instead use Code: `println $(< file)` as it is more efficient. This is because bash has file opening and reading the file instead of creating a new process to do this. (Don't worry you'll get to creating and destroying processes and switching between them later in the semester.)

G.2 Thanks for sharing this Dan. I need to get a few things fixed in my code before I give it a go, but I'm glad I have this. _______________ -m

G.3 We tried gdb, we tried make clean and make, and yet... It seems to be breaking with the bit in Player.hpp about Code: gc->toward_and_times.push_back(toward_and_space); I've not seen anything so bizarre in recent memory. [...] I may be getting ridiculously incredibly lucky. Hope this helps.

G.4 I was hoping to hear some chatter on this project. Chris, I can look at this more closely when I return, but what I think is going on is that you need to make sure your Python files get packaged up and sent to the slave machines. [...] Edited by Mike C.: Please see my note below about getting the file from S3.

G.4.1 My fault. I fixed the permissions, and now you can get it. Remember to use the copy on S3 from EC2 because it's faster.

G.4.2 Mike Colagrosso wrote: One thing that I've done is I've replicated that first line (the movie ID) onto the end of every other line in the file. [...] Hadoop is good and preprocessing the data for you, but my modification was so simple that I did it in the shell.
## Zooming on OS Proje cts 08

<table>
<thead>
<tr>
<th>G.4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I was hoping to hear some chatter on this project. Chris, I can look at this more closely when I return, but what I think is going on is that you need to make sure your Python files get packaged up and sent to the slave machines.</td>
<td></td>
</tr>
<tr>
<td>[... Edit by Mike C.: Please see my note below about getting the file from S3.</td>
<td></td>
</tr>
<tr>
<td>G.4.1</td>
<td></td>
</tr>
<tr>
<td>My fault. I fixed the permissions, and now you can get it. Remember to use the copy on S3 from EC2 because it's faster.</td>
<td></td>
</tr>
<tr>
<td>G.4.2</td>
<td></td>
</tr>
<tr>
<td>Mike Colagrosso wrote: One thing that I've done is I've replicated that first line (the movie ID) onto the end of every other line in the file. [...]Hadop is good and preprocessing the data for you, but my modification was so simple that I did it in the shell.</td>
<td></td>
</tr>
<tr>
<td>Were you able to sort this out? [... Is this correct?</td>
<td></td>
</tr>
</tbody>
</table>
Expanding a cell

G.4.2 Mike Colagrosso wrote: One thing that I've done is I've replicated that first line (the movie ID) onto the end of every other line in the file. [....] Hadoop is good and preprocessing the data for you, but my modification was so simple that I did it in the shell.

Were you able to sort this out? I run the same example in the Project 4 handout Code: [root@hadoop-0.16.1]# ./bin/hadoop jar hadoop-0.16.1-examples.jar pi 10 10000000 Does the web interface startup on port 50030? I have been unable to set this. I spoke with a couple teams, who all missed out on this experience my team is having. The web interface does start up. Daniel and I have been able to really point our fingers at DFS. Running ./bin/hadoop dfsadmin -report on the EC2 machines show that their are 0 datanodes. Again, any help would be appreciated. (We have started on our scripts, but are only hoping we'll be able to plug them in easily.) Thanks. __________________ -m If it's still useful... Code: 

```
#!/bin/bash
for i in $1/* do
  NUM=`grep ':' $i | sed -r 's/\([[:digit:]]\)+\:\:\:\*:/*/g'
  sed '1d' $i > tmp.out
  sed -r "s/\([^[:space:]]*\)\([^[:space:]]*\)/\1\2/g" tmp.out > $i
done
```

Code: 

```
#!/bin/bash
for i in $1/* do
  NUM=`grep ':' $i | sed -r 's/\([[:digit:]]\)+\:\:\:\*:/*/g'
  sed '1d' $i > tmp.out
  sed -r "s/\([^[:space:]]*\)\([^[:space:]]*\)/\1\2/g" tmp.out > $i
done
```

Added the echo at the bottom of the for loop for the impatient on those 17k file sets (me). Yes it slows it down some but meh. I'd rather know where I'm at. Is there a way to guarantee that one reducer will more or less see all the key/values? Or at least the summary results from them? I'd like to be able to sort on summary statistics, and maybe print, for example, the top-10-(blank)est movies. Anyone have ideas? I would do this by chaining two MapReduce jobs together. The original MapReduce paper from Google said that they do this often. I just want to make sure that we are to hand in our team evals the day of the final, and we are not charged slip days for this, since they are physical papers, and not a submission through BB. Is this correct? [x]
Agora

• No need for dictionary
• No costly description and storage of all possible formats, labels etc...
• Exploits *differences* in layout, labels and punctuation distribution
• Results reflect meaningful turns in collective discussion
Evolution in time

When does a collective discussion get momentum?
Parsing on the fly

• Forum in Computer Science
• OS Projects 1st semester 08
  – 53 threads in a forum and 166 posts
After 1 week

- Tuning not performed yet

Example:

I know it's important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? [...] So far it hasn't been a problem per se, but just disconcerting.

Your shell is not using true regular expressions. [...] I hooked these updates into the conference website. MobiSys 08 There's many more worth suggesting, including at least one account from one of our classmates.
After 2 weeks

• Tuning achieved

Example
After 6 weeks

- Six moments in discussion proper

<table>
<thead>
<tr>
<th>G</th>
<th>I know it's important to check the options and arguments to your script, but is it also common practice to check arguments to functions you define? My guess is yes, but it seems to make my code a little bloated sometimes. For instance, if you have a function that takes a directory as an argument, not only would you have to check that the number of arguments is correct, but also that the first argument really is a directory. [...] Wouldn't we want the html to update along with the new rss feeds? Right now I'm calling aggregate at the end of fetch and then cleaning up afterward. What have you guys done about this little discontinuity?</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.1</td>
<td>For the MAILTO part, will &quot;challen&quot; automatically send to my Mines address or do I need to specify &quot;@mines.edu&quot; or does either work? [...]bump bump?</td>
</tr>
<tr>
<td>G.2</td>
<td>so i'm trying to delete an entire directory with contents in it, yet rd doesn't work in alamode. [...]If you're sneaky enough you can sort the original format with one command. Here is a link to get you started on the road to sneaky</td>
</tr>
<tr>
<td>G.3</td>
<td>I guess by my robustness comment, I was thinking about the one pipe or one file redirect thing. [...]My question was whether the command is supposed to be used with both a value and a &lt;command&gt; or not.</td>
</tr>
<tr>
<td>G.4</td>
<td>max [is back] wrote: With our variable expansion, do we need to support things like Code: echo $DIR/$FILE It would be cool, but no, it's not necessary. [...] That's what my shell does.</td>
</tr>
<tr>
<td>G.5</td>
<td>Michael, yes they are. (And, yes you can.) [...]Is it okay if in our shells we simply print an error message when there's no closing quote?</td>
</tr>
<tr>
<td>G.6</td>
<td>I've been researching execve and it seems to me that you must include a file for exec to get it's new code from, as well as an array of arguments that you want the new process to have. [...] We can't exactly exec them...</td>
</tr>
</tbody>
</table>
After 14 weeks: end of term

• 4 moments: re-arranged

D.1 Dan and I worked on some to convert the dates in the RSS feeds to something that is easy to sort: Code: $ date -d "Fri, 11 Jan 2008 19:49:49 +0000" +%%y-%%d-%H-%%M-%S" 08-01-11-12-49-49 and here’s a version that works on Mac OS X: Code: $ date -j -f "%%a, %%d %b %C%%y %%H:%%M:%%S %z" "Fri, 11 Jan 2008 19:49:49 +0000" +%%y-%%m-%%d-%H-%%M-%%S" 08-01-11-12-49-49 so I’m trying to delete an entire directory with contents in it, yet it doesn’t work in alamode. On the contrary, everything inside of the $( ) is interpreted normally just as if it were its own command. (Also, it is possible to jump from the beginning to the end or vice-versa of one of these blocks in vim using % while that’s not possible with the backquotes.) On a side note, the section that details this also suggests that if you need to do Code: $ blah $(cat file) you instead use Code: $ blah < file as it is more efficient. This is because bash handles opening and reading the file it instead of creating a new process to do this. (Don’t worry you’ll get to creating and destroying processes and switching between them later in the semester.)

G.2 Thanks for sharing this Dan. I need to get a few things fixed in my code before I give it a go, but I’m glad I have this. ____________________ -m

G.3 We tried gdb, we tried make clean and make, and yet... It seems to be breaking with the bit in Player.hpp about: Code: gc->turnaround_times.push_back(turnaround); I’ve not seen anything so bizarre in recent memory. [...] I may be getting ridiculously incredibly lucky. Hope this helps.

G.4 I was hoping to hear some chatter on this project. Chris, I can look at this more closely when I return, but what I think is going on is that you need to make sure your Python files get packaged up and sent to the slave machines. [...] Edited by Mike C.: Please see my note below about getting the file from S3.
Interpretation

• Detected higher level pattern moment G1
• Code exchange and collaboration between students
Summing up

- Agora helps monitoring students’ discussion
  - Works on text
    - gives access to content
  - On line

- Agora is robust
  - Does not need external resources

- Agora is adaptive
  - Domain-free
  - Multilingual
  - Processes discussion lists as well
but

• Visualisation is too coarse
  – Give number of masked items
    • [8 posts...] instead of [...]  
  – Give duration of main functional segments

• Give access to more significant text
  – It is difficult to get an idea of the current discussion through snippets
Further work

• Tests on different formats
• Test more languages
• Large on-line discussions
  – Monitoring virtual classes on many tasks
• Visualisation
  – Provide options
Thank you
<thread name="OS Projects">
  <message id="155">
    <header>
      <author>Mike Colagrosso</author>
      <subject>Code snippet from sed discussion</subject>
    </header>
    <body>
      <td class="postbody">
        <table width="90%" cellspacing="1" cellpadding="3" class="code" align="center">
          <tr>
            <td class="row1"><span class="genmed"><b>Code:</b></span></td>
          </tr>
          <tr>
            <td class="row2"><span class="postbody"><font color="#006600">cat index.xml | grep enclosure | sed 's/^.*url="\([^"\]*\)".*$/\1'</font></span></td>
          </tr>
        </table>
      </td>
    </body>
  </message>

  <message id="156">
    <header>
      <msgref id="155"/>
      <author>AndyMan1</author>
      <datetime>16/09/2007 23:15</datetime>
      <subject></subject>
    </header>
    <body>
      <span class="postbody">I found this cool list of sed one-liners ( *mimes a cigar a la Groucho*).<br />
      It has examples of doing all sorts of short commands with sed like double spacing a file, deleting every 8th line, print only lines that don't match regexp, etc.<br />
      Nothing in it seemed to be too revealing in terms of our project. It has a few examples that might be useful as a starting point.<br />
      <a href="http://sed.sourceforge.net/sed1line.txt" target="_blank">http://sed.sourceforge.net/sed1line.txt</a></span>
    </body>
  </message>
</thread>
Mike Colagrosso wrote:

Code snippet from sed discussion

```
cat index.xml | grep enclosure | sed 's/^.*url="(\[^"\]*\)".*$/\1/'
```

AndyMan1 replied:

I found this cool list of sed one-liners ( *mimes a cigar a la Groucho*).

It has examples of doing all sorts of short commands with sed like double spacing a file, deleting every 8th line, print only lines that don't match regexp, etc.

Nothing in it seemed to be too revealing in terms of our project. It has a few examples that might be useful as a starting point.
Algorithm

- Detect background
- Process unit
  - Calculate rank
  - Divide
    - Detect breaks
    - Set borders
    - Group similar
      - Set wrappers
  - Get wrapped sub-unit
Find a new set of features

• Disappearance of common items
  – Greetings
  – Images
  – ...

• Appearance of new items
  – Quotes from other messages
  – Images
  – Code (for computer sciences)
  – ...

Titre

Agora
Bonjour, je cherche une traduction de la RFC pour http1.1 (je n'ai trouvé qu'une vieille 'version'). Ai commencé à la lire; mais 163 pages, c'est vraiment beaucoup! Alors si quelqu'un connaît un site c'est vraiment bienvenu... sauf, peut être tu trouveras ton bonheur chez : http://abdcfr.free.fr/ Joel Merci pour ta réponse, j'y suis déjà allé, mais le RFC 2060 (http1.1) n'y est pas. C'est là que j'ai trouvé la traduction de la première version http... [___] Bonjour, sincères remerciements à Pascal pour VB (parce que tu le veux bien... et je te remercie çà !). Pour ceux qui ont à la bourre comme moi, ya un site pas mal à voir sur java: <a href="http://perso.wanadoo.fr/jm.goudoux/java/tutorial/indexacritrames.htm">http://perso.wanadoo.fr/jm.goudoux/java/tutorial/indexacritrames.htm</a>.


G.1 Copie de l'actualité publiée le 07/11/03 : Bonjour, J'ai bien reçu VB. Merci encore. Kc

G.2 Merci j'vais être la seule...

G.3 En ce que me concerne, je pense passer ce module en tout dernier lieu. [___] le lien pour vous envoyer un email à Monsieur Sidir renvoie une erreur.

G.4 Merci mais apparemment le devoir n°1 de ce module n'est plus disponible ? (lorsque je clique sur le lien, j'abouti sur le cours ... [___] Bon courage....

G.5 Hello, Pour les utilisateurs windows, voici quelques infos pour bien débuter avec easy php. - D'abord installer le programme (:) qui se trouve sur l'adresse plus bas - Le lancer, puis cliquer sur le bouton droit sur son icône dans la barre des tâches et choisir 'configuration' - easyphp - Vérifier que les serveurs sont au vert - Pour exécuter les scripts, placer la page web dans le répertoire www (ex: C:\Program Files\EasyPHP1.7\www) - Ouvrir la page (qui doit avoir l'extension .PHP) dans le navigateur avec l'adresse: http://127.0.0.1/mage/php ou encore http://localhost/mage/php [___] Pour php, allez visiter ce site


G.7 Bonjour !! [___] Décidé il y a JSPMyAdmin mais il ne fonctionne pas, quelqu'un sait-il comment faire ?

G.8 Bonjour, j'ai changé d'ordinateur et j'ai donc du réinstaller tomcat etc. [___] Peut-être que cela marche maintenant...

G.9 Si l'exposant avait été égal à 2, la manche resterait la même mais 0.11010 * par 10 puissance -2 donne 0.0011010. (Il y a un zéro de plus après la virgule.)

G.10 Pour l'avoir testé sur un essai et avec suprise aussi : cela efface toute l'appli web... et pas de sauvegarde... C'est vrai que c'est trompeur, retirer=>supprimer, aussi pour moi. [___] Bonjour,

G.11 Deja que Monsieur Délage ne veut pas nous donner son cours dans un format imprimable (pourtant nous payons - son éditeur devrait pouvoir nous accorder cela) je me demande bien comment je vais y arriver avec sa partie...

G.12 Et maintenant on peut télécharger le cours - super - je retire ce que j'ai dit l'autre jour ! [___] Merci.

G.13 J'ai beau cherché une documentation sur l'assembleur mips r3000, je n'ai rien trouvé de clair. Quelqu'un aurait-il une idée pour trouver une doc claire avec la description de toutes les instructions du r3000 ? [___] pour s'entraîner !

G.14 Bonjour, [___] L_R.

G.15 Sous quelle forme faut-il rendre le projet ? [___] promis de particulier à faire à ce niveau si ce n'est indiquer le lien vers votre site et continuer à discuter si vous le souhaitez

G.16 Bonjour, je suis nouveau et j'ai commencé les cours et exercices. [___] merci Je crois que Mr Cérin en recommande 2, de Patterson et Hennessy en anglais-Lequel serait le plus utile?

G.17 Bonjour, j'ai été très occupé depuis septembre a cause de mon travail et pour cette session je voudrais préparer l option 1 avec Diana et éventuellement Richard (il faut qu'il voit). [___] L'archive en question se trouve comme d'habitude dans la rubrique tutorat : regroupements m2

G.18 Bonjour à tous, [___] Cordialement, E. Marty
G.1  
Taken/interpreted from Webopedia (http://www.webopedia.com/Term/b/big_endian.html)... Very nice.

jemman Jeremy Norman [___]Alex

G.2  
What is the purpose of System Calls and how do system calls relate to the operating system and to the concept of dual-mode (kernel mode and user mode) operations? System calls allow user-level processes to request services from the operating system that the process itself is not allowed to do. These system calls are classified into broad six categories - Filesystem, Process, Scheduling, Interprocess Communication, Networking, and Miscellaneous. [___] Code: du -s disk usage by you (assuming you are in your home directory) Host name: Code: uname -n OS: Code: uname -o

G.2.1  
IP Address Code: /sbin/ifconfig ________________ Keith Mitchell "Fools! [___]lt's made by lazy man trying to find easier ways to do something." - Robert Heinlein

G.2.2  
Quote: '[Chevenot Hall] is open from before 8am until 6pm on weekdays.' - MACS Secretary The Alamode Lab is open whenever the building is open, as well as much later into the night. [___]BBCode needs a paraphrase option...

G.2.3  
Remember that we're meeting in the Alamode Lab today to work on Project 1. [___]So, I double click on my program, it gets its process and is ready to do some work

G.2.4  
The short term scheduler runs after every millisecond or whatever (the amount of time between time interrupts), so it definitely is executed the most. [___] Theirs know thyme when thee presence.

G.2.5  
Good explanation, Squid. [___] If your Project 1 code is still working, it will download them automatically. (Also, this post will show up in your webpage).

G.2.6  
Here is another situation to consider: [___] Use the homework assignment to study for the exam, and doing it nets you very easy points.

G.3  
Let's try a different way to explain this one. [___] Later, Tom Brokaw shows up and enjoys two wonderful meals.

From in class, we want three processes to each print out "I am process NUMBER" in the order F1, P2, P3 repeating the order in an infinite loop. [___] I don't know if P2 would eventually lead to deadlock, and I'm too tired to figure it out right now. [___] Brandon Cruz Ardiente Like R-D-N-I-T

G.4  
Brandon Ardiente wrote: After the text on Tuesday and tonight's programming project, I'm definitely feeling devious. [___] I'd like to say no, but I don't know the reasons why not.

G.4.1  
The disadvantage of implementing synchronization with special machine instructions is that they are hardware-specific, and some machines might not have them available. [___] Repeat [call Tom]; raise Lope; Sign = Nicole; while((Code is up) AND (Sign = Nicole)) skip; [eat dinner] lower Lope; forever;

G.4.2  
I connected to Raspberry through ssh, made a new directory in my home directory called "boost", then copied all of the files from /home/mcolagros/boost-examples/ into my directory. [___] How many bits are needed to represent the page number of a process? 2 bits

G.5  
Critique of paging: [___] Yeah, I also added the export line to my .bashrc (and ran it in the terminal) before I tried make and I got the same results as Brandon. ________________ Sara McF

Sara, [___] Translating logical to physical addresses: Code: 00110 (6) -> 010110 11010 (26) -> 011010 10011 (19) -> 000011 01101 (13) -> 101101

G.6  
Teresa Davies wrote: I've found that when I ssh into alamode, I have to set LD_LIBRARY_PATH every time. [___] You'll use the readline and Boost threading libraries in this project.

G.6.1  
I was reading question 6 and I wasn't sure exactly what a process flow graph entailed. [___] Remember the problem where I asked you to add semaphores to the code to ensure it printed;

G.6.2  
We've always had a distinction between page number and offset. [___] If the resource you are trying to create exclusion on is actually an I/O device, you end up resorting to polling the device and wasting CPU cycles that could be used for actual work.
**G.3**

Wow. That's an impressive splash! You open a file, fork, and in the child `dup` the open file's file descriptor onto standard in, the call exec. There aren't any pipes involved.

**G.3.1**

In running through the tests listed in the tests post, I ran this test (from Sara): [...]Do we need to capture this situation properly, or can I call my bug a feature?

**G.3.2**

Mine is super fancy... [...]I most certainly will be fishing from now on ________________

: Like R-D-N-T

**G.3.3**

Or, you could scope lock on the cout statement... [...]Hopefully that helps a little bit.

____________________ -Sara McF

I don't think you really want to specifically address a thread or directly tell a thread to do a specific entry (like the first thread doing the third entry). [...]In my example, you can see that two seconds after the thread first accesses the request is when it starts to service the request.
As of right now, I have active waiting for each thread. I can slow this down by waiting for a few seconds before polling...however...this still doesn't help that much.

Suggestions?

nevermind, figured it out.

What was your solution? If you figure out your own question, it's always great to answer yourself in case someone else runs into the same issue.

I just kicked it and it started to work. That usually solves a lot of coding problems.

EDIT:
Ok, now that I have a little more time I'll add to my smart ass answer.

Basically I called a wait on a scopedlock within each thread when the thread had no work to do. Then after an item is pushed onto the priority queue notify_all is called. This needs some tweaking that I'll leave to the rest of you.
Comparison with activity graph

Discussion
Start + 4 weeks

- Three moments in discussion proper