

クラウド・テクノロジー研究部会  
**セマンティックWebプログラミング  
ことはじめ**

2012/08/30

キャノンソフト情報システム

上村準也

Build Flexible Applications with Graph Data

## セマンティックWeb プログラミング



O'REILLY®  
オライリー・ジャパン

Toby Segaran, Colin Evans, Jamie Taylor 著  
大向 一輝、加藤 文彦、中尾 光輝、山本 泰智 監訳  
玉川 竜司 訳

こちらの書籍を元に  
3つ組みデータの考え方と  
プログラミングについて  
簡単な説明をします

4部構成の書籍ですが  
第1部の始めにあたる話です

2010年06月 発行  
312ページ

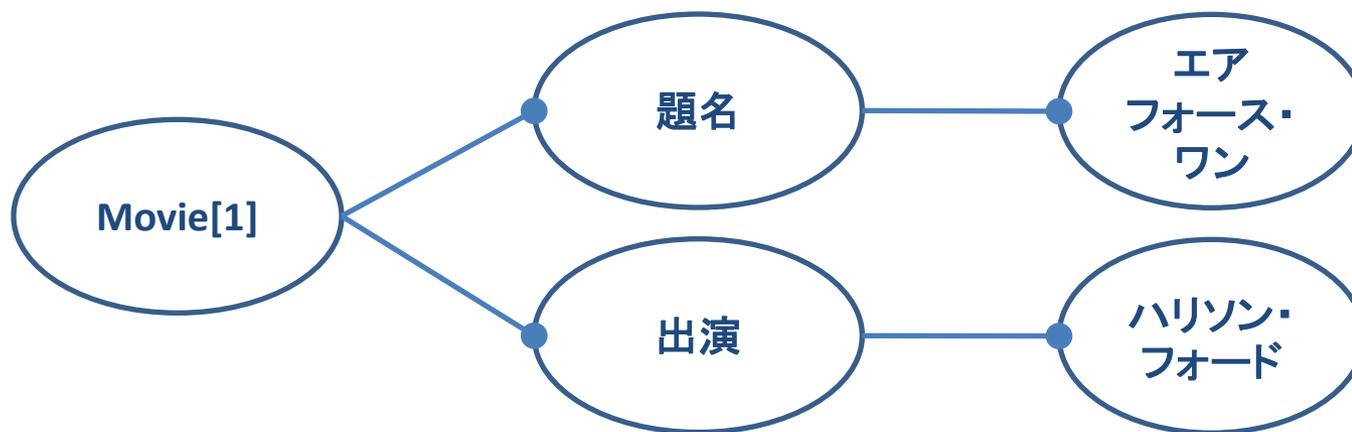
<https://www.oreilly.co.jp/books/9784873114521/>

# 3つ組データの考え方

- 特に前提ない、主語、述語、目的語に直すだけ



- 例えば映画なら...



# 3つ組データの考え方

- 簡単にCSV形式のテキストデータとして見てみる
- 主語はデータにとってユニークになるよう工夫

	A	B	C
1	主語 (Subject)	述語 (Predicate)	目的語 (Object)
108813	/authority/imdb/title/tt0046534	name	The War of the Worlds
108814	/authority/imdb/title/tt0046534	starring	/en/gene_barry
108815	/authority/imdb/title/tt0046534	starring	/guid/9202a8c04000641f8000000006be005d
108816	/authority/imdb/title/tt0046534	starring	/en/ann_robinson
108817	/authority/imdb/title/tt0046534	directed_by	/en/byron_haskin
108818	/authority/imdb/title/tt0118571	name	Air Force One
108819	/authority/imdb/title/tt0118571	starring	/guid/9202a8c04000641f80000000055cc5f0
108820	/authority/imdb/title/tt0118571	starring	/en/liesel_matthews
108821	/authority/imdb/title/tt0118571	starring	/en/xander_berkeley
108822	/authority/imdb/title/tt0118571	starring	/en/harrison_ford
108823	/authority/imdb/title/tt0118571	starring	/en/paul_guilfoyle
108824	/authority/imdb/title/tt0118571	starring	/en/glenn_close
108825	/authority/imdb/title/tt0118571	starring	/en/william_h_macy
108826	/authority/imdb/title/tt0118571	starring	/en/wendy_crewson
108827	/authority/imdb/title/tt0118571	starring	/en/gary_oldman
108828	/authority/imdb/title/tt0118571	directed_by	/en/wolfgang_petersen
108829	/en/emily_procter	name	Emily Procter

# 3つ組データの考え方

- グラフなので、主語を目的語にしてつなげる
- さらに別の情報を関連して記憶できる

	A	B	C
1	主語 (Subject)	述語 (Predicate)	目的語 (Object)
26616	/authority/imdb/title/tt0046834	directed_by	/guid/9202a8c04000641f8000000000d41069
26617	<b>/en/harrison_ford</b>	<b>name</b>	<b>Harrison Ford</b>
108817	/authority/imdb/title/tt0046534	directed_by	/en/byron_haskin
108818	/authority/imdb/title/tt0118571	name	Air Force One
108819	/authority/imdb/title/tt0118571	starring	/guid/9202a8c04000641f800000000055cc5f0
108820	/authority/imdb/title/tt0118571	starring	/en/liesel_matthews
108821	/authority/imdb/title/tt0118571	starring	/en/xander_berkeley
108822	<b>/authority/imdb/title/tt0118571</b>	<b>starring</b>	<b>/en/harrison_ford</b>
108823	/authority/imdb/title/tt0118571	starring	/en/paul_guilfoyle
108824	/authority/imdb/title/tt0118571	starring	/en/glenn_close
108825	/authority/imdb/title/tt0118571	starring	/en/william_h_macy
108826	/authority/imdb/title/tt0118571	starring	/en/wendy_crewson
108827	/authority/imdb/title/tt0118571	starring	/en/gary_oldman
108828	<b>/authority/imdb/title/tt0118571</b>	<b>directed_by</b>	<b>/en/wolfgang_petersen</b>
108829	/en/emily_procter	name	Emily Procter
129919	<b>/en/wolfgang_petersen</b>	<b>name</b>	<b>Wolfgang Petersen</b>
129920	/guid/9202a8c04000641f8000000000	name	Randy Lee

# 3つ組データの考え方

- 種類の異なるデータをどんどん追加できる

	A	B	C
1	主語 (Subject)	述語 (Predicate)	目的語 (Object)
26616	/authority/imdb/title/tt0046834	directed_by	/guid/9202a8c04000641f8000000000d41069
26617	/en/harrison_ford	name	Harrison Ford
108817	/authority/imdb/title/tt0046534	directed_by	/en/byron_haskin
108818	/authority/imdb/title/tt0118571	name	Air Force One
108819	/authority/imdb/title/tt0118571	starring	/guid/9202a8c04000641f800000000055cc5f0
108820	/authority/imdb/title/tt0118571	starring	/en/liesel_matthews
108821	/authority/imdb/title/tt0118571	starring	/en/xander_berkeley
108822	/authority/imdb/title/tt0118571	starring	/en/harrison_ford
108823	/authority/imdb/title/tt0118571	starring	/en/paul_guilfoyle
108824	/authority/imdb/title/tt0118571	starring	/en/glenn_close
108825	/authority/imdb/title/tt0118571	starring	/en/william_h_macy
108826	/authority/imdb/title/tt0118571	starring	/en/wendy_crewson
108827	/authority/imdb/title/tt0118571	starring	/en/gary_oldman
108828	/authority/imdb/title/tt0118571	directed_by	/en/wolfgang_petersen
108829	/en/emily_procter	name	Emily Procter
129919	/en/wolfgang_petersen	name	Wolfgang Petersen
129920	/guid/9202a8c04000641f8000000000	name	Randy Lee
185037	/en/rita_wolf	name	Rita Wolf
185038	/authority/imdb/title/tt0118571	written_by	/en/andrew_w_marlowe
185039	/en/andrew_w_marlowe	name	Andrew W. Marlowe

# 3つ組データの検索

- 単純な検索はもちろん可能
- 主語の形式が異なるが...
  - 題名が述語:starringで記録されていれば...
  - 異なるデータソースから集めたものも

	A	B	C
1	主語 (Subject)	述語 (Predicate)	目的語 (Object)
36451	/authority/imdb/title/tt0066601	starring	/en/harrison_ford
36452	/authority/imdb/title/tt0083866	starring	/en/harrison_ford
36468	/authority/netflix/movie/70077535	starring	/en/harrison_ford
36469	/authority/netflix/movie/70082398	starring	/en/harrison_ford
36475	/en/indiana_jones_4	starring	/en/harrison_ford
36476	/en/indiana_jones_and_the_last_crusade	starring	/en/harrison_ford
36477	/en/indiana_jones_and_the_temple_of_doom	starring	/en/harrison_ford
36478	/en/raiders_of_the_lost_ark	starring	/en/harrison_ford
36490	/guid/9202a8c04000641f800000000552fdf4	starring	/en/harrison_ford
36491	/guid/9202a8c04000641f80000000057a001d	starring	/en/harrison_ford
36492	/guid/9202a8c04000641f8000000005aa9682	starring	/en/harrison_ford

# 3つ組データの検索

- パターンの検索が可能
- 主語(S)、述語(P)、目的語(O)の形式で考えれば...

- S=映画 P=directed\_by O=俳優 かつ
- S=映画 P=starring O=俳優 であれば

- 監督 = 出演者である映画と俳優をあらわす

# 3つ組データの検索

- 「3.1.2クエリ言語の実装」で紹介されているサンプルをGroovyで実装し直してみた

The screenshot shows a web browser window displaying a GitHub Gist page. The browser's address bar shows the URL `https://gist.github.com/2904730`. The page header includes the GitHub logo and the user name `uemuraj`. The main content area shows the gist details for ID `2904730`, including a description in Japanese: "セマンティックWebプログラミングの第2章から第3章の内容を試す。" (Try the content from chapter 2 to chapter 3 of Semantic Web Programming). Below the description, there are buttons for `edit`, `download`, and `star`. The gist content is a Groovy file named `SimpleGraph.groovy`, which defines a `SimpleGraph` class with three lists: `spo`, `pos`, and `osp`. The code is as follows:

```
SimpleGraph.groovy # embed raw
1 //
2 // セマンティックWebプログラミングの第2章から第3章の内容を試す。
3 //
4
5 class SimpleGraph {
6
7     // 各インデックスとも [ a : [ b : [ c, .. ] ] ] の様な格納方法
8     def spo = [:]
9     def pos = [:]
10    def osp = [:]
```

On the right side of the page, there is a section for `Revisions` showing three revisions by the user `uemuraj` (owner):

- Revision `01e165` (4 minutes ago)
- Revision `12d208` (5 minutes ago)
- Revision `8feef83` (3 months ago)

# 3つ組データの検索

?movie,directed\_by,?actor

?movie,starring,?actor

The screenshot shows a Jenkins web interface. The browser address bar displays `jenkins.sierra1337.info/job/simple-graph/3/consoleFull`. The page title is "Jenkins" and the user is logged in as "uemuraj". The breadcrumb navigation shows "Jenkins > simple-graph > #3". On the left sidebar, there are several navigation links: "プロジェクトへ戻る", "状態", "変更履歴", "コンソール出力", "説明の編集", "ビルドを削除", "パラメータ", "タグ設定", and "前のビルド". The main content area is titled "コンソール出力" and displays the following console output:

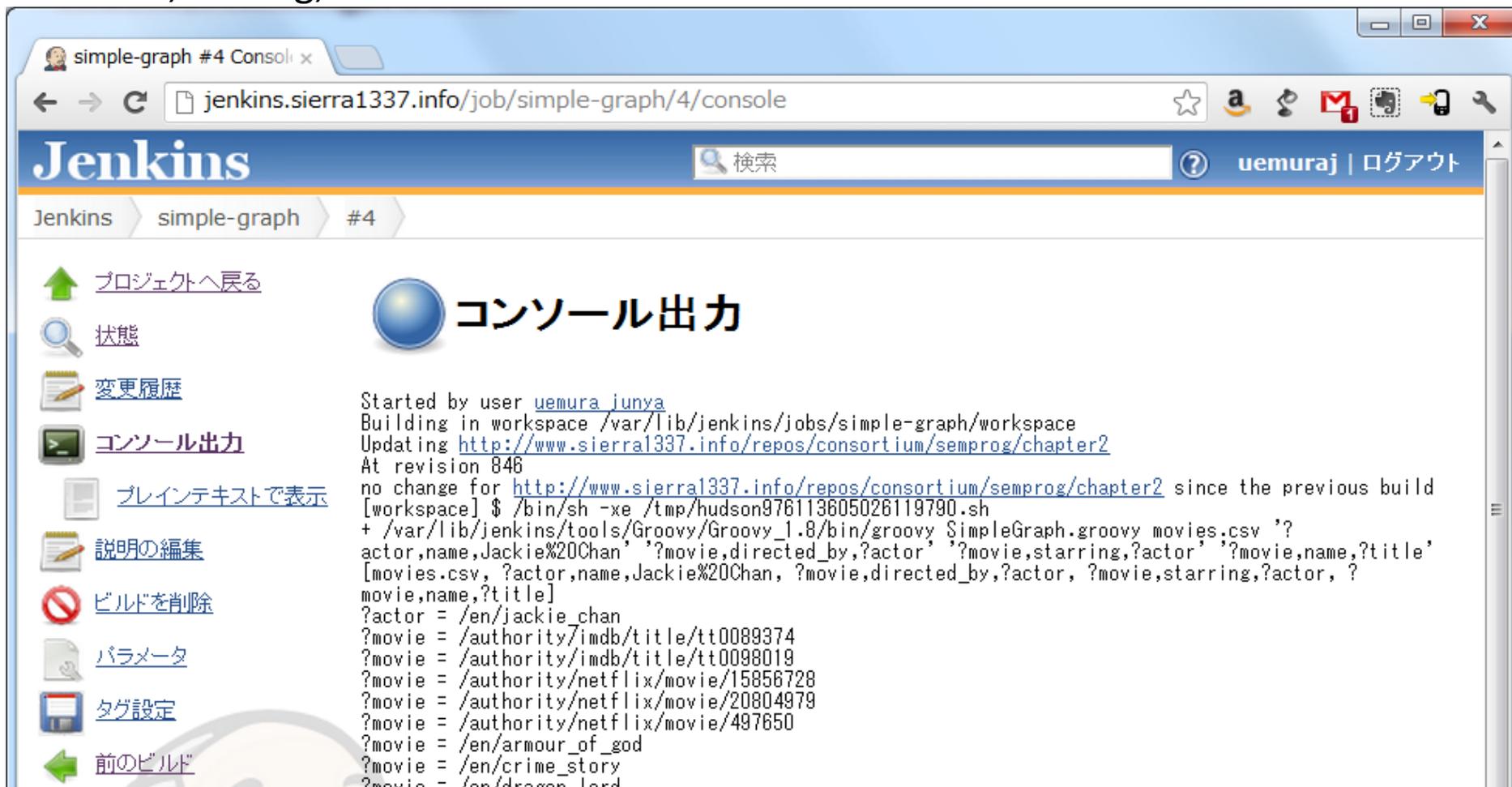
```
Started by user uemura_junya
Building in workspace /var/lib/jenkins/jobs/simple-graph/workspace
Updating http://www.sierra1337.info/repos/consortium/semprog/chapter2
At revision 846
no change for http://www.sierra1337.info/repos/consortium/semprog/chapter2 since the previous build
[workspace] $ /bin/sh -xe /tmp/hudson1599737837702111026.sh
+ /var/lib/jenkins/tools/Groovy/Groovy_1.8/bin/groovy SimpleGraph.groovy movies.csv '?
movie,directed_by,?actor' '?movie,starring,?actor' '?movie,name,?title'
[movies.csv, ?movie,directed_by,?actor, ?movie,starring,?actor, ?movie,name,?title]
?movie = /authority/imdb/title/tt0000582
?movie = /authority/imdb/title/tt0002130
?movie = /authority/imdb/title/tt0003805
?movie = /authority/imdb/title/tt0004026
?movie = /authority/imdb/title/tt0005044
?movie = /authority/imdb/title/tt0005074
?movie = /authority/imdb/title/tt0005810
?movie = /authority/imdb/title/tt0008177
?movie = /authority/imdb/title/tt0006548
?movie = /authority/imdb/title/tt0007813
```

# 3つ組データの検索

?actor,name,Jackie%20Chan

?movie,directed\_by,?actor

?movie,starring,?actor



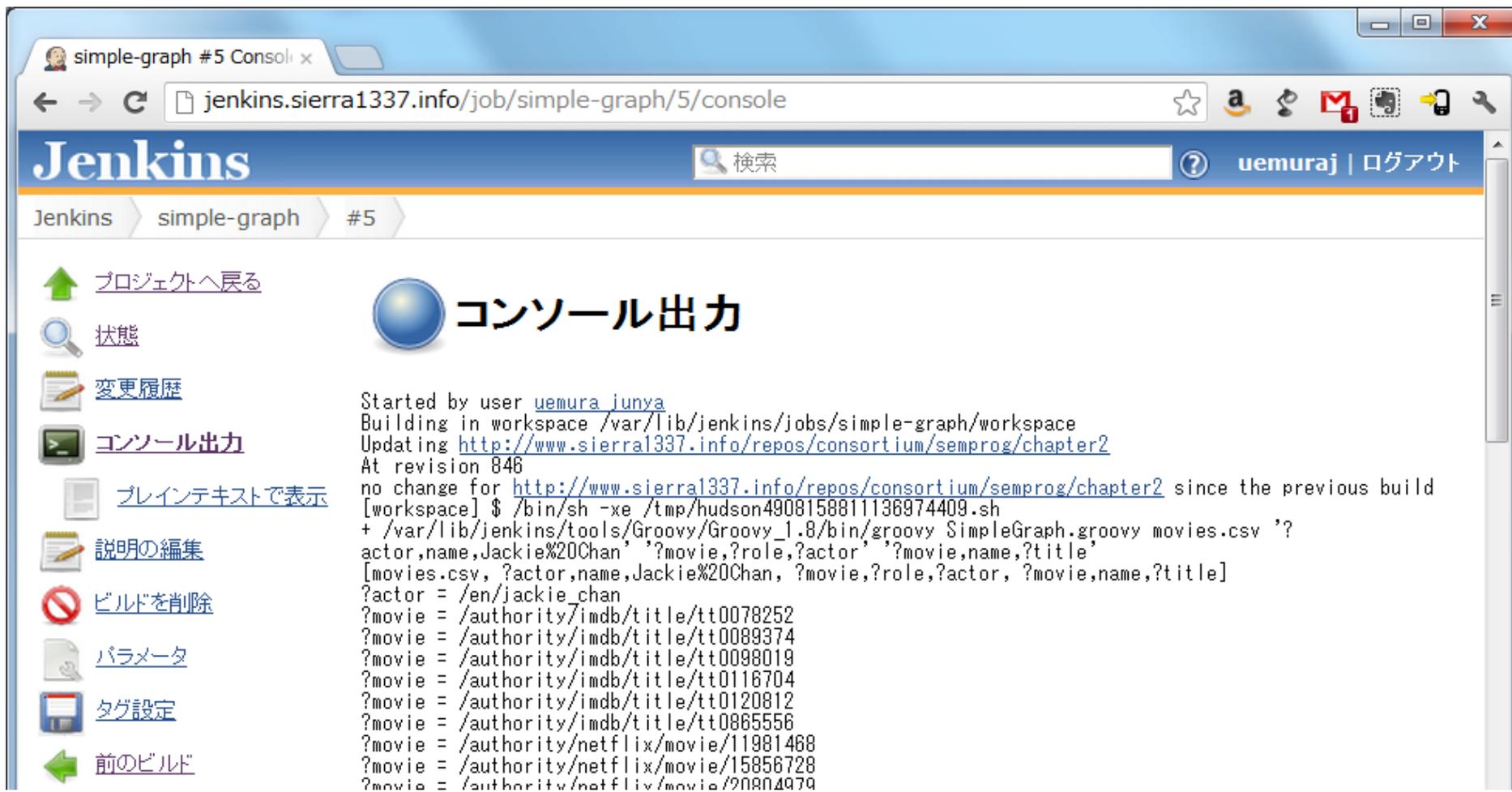
The screenshot shows the Jenkins web interface for a job named 'simple-graph #4'. The console output displays the following text:

```
Started by user uemura_junya
Building in workspace /var/lib/jenkins/jobs/simple-graph/workspace
Updating http://www.sierra1337.info/repos/consortium/semprog/chapter2
At revision 846
no change for http://www.sierra1337.info/repos/consortium/semprog/chapter2 since the previous build
[workspace] $ /bin/sh -xe /tmp/hudson976113605026119790.sh
+ /var/lib/jenkins/tools/Groovy/Groovy_1.8/bin/groovy SimpleGraph.groovy movies.csv '?
actor,name,Jackie%20Chan' '?movie,directed_by,?actor' '?movie,starring,?actor' '?movie,name,?title'
[movies.csv, ?actor,name,Jackie%20Chan, ?movie,directed_by,?actor, ?movie,starring,?actor, ?
movie,name,?title]
?actor = /en/jackie_chan
?movie = /authority/imdb/title/tt0089374
?movie = /authority/imdb/title/tt0098019
?movie = /authority/netflix/movie/15856728
?movie = /authority/netflix/movie/20804979
?movie = /authority/netflix/movie/497850
?movie = /en/armour_of_god
?movie = /en/crime_story
?movie = /en/dragon_lord
```

# 3つ組データの検索

?actor,name,Jackie%20Chan

?movie,?role,?actor



The screenshot shows the Jenkins web interface for job 'simple-graph #5'. The console output displays the following text:

```
Started by user uemura_junya
Building in workspace /var/lib/jenkins/jobs/simple-graph/workspace
Updating http://www.sierra1337.info/repos/consortium/semprog/chapter2
At revision 846
no change for http://www.sierra1337.info/repos/consortium/semprog/chapter2 since the previous build
[workspace] $ /bin/sh -xe /tmp/hudson4908158811136974409.sh
+ /var/lib/jenkins/tools/Groovy/Groovy_1.8/bin/groovy SimpleGraph.groovy movies.csv '?
actor,name,Jackie%20Chan' '?movie,?role,?actor' '?movie,name,?title'
[movies.csv, ?actor,name,Jackie%20Chan, ?movie,?role,?actor, ?movie,name,?title]
?actor = /en/jackie_chan
?movie = /authority/imdb/title/tt0078252
?movie = /authority/imdb/title/tt0089374
?movie = /authority/imdb/title/tt0098019
?movie = /authority/imdb/title/tt0116704
?movie = /authority/imdb/title/tt0120812
?movie = /authority/imdb/title/tt0865556
?movie = /authority/netflix/movie/11981468
?movie = /authority/netflix/movie/15856728
?movie = /authority/netflix/movie/208804979
```

# 続きは...

- DBpedia [[ja.dbpedia.org](http://ja.dbpedia.org)] へ行ってみてください
  - もっと面白いデータをSPARQLで検索できます
- 「セマンティックWebプログラミング」
  - まだまだ続きがあります、読んでみてください
- とにかくたくさんの3つ組データを持ちたい
  - いわゆるSNS系データを処理したいとすると...
  - 何をベースに記憶するか、どう検索するかは 研究 & 調査の余地 がありそう
- さらに踏み込んだ次の発表をお聞きください