

Data

22/07/2009

Previous 1/1 Next

- Common
- Music
- News
- Product

Channel

XML

Explorer Category Search

- Common
- Addresses
- Thomasnet
- Zipcodedatabase

Channel Name(*): History Version

Home Page(*):

Crawled Link Pattern:

Sample Page:

Link Pattern:

Extract Type: Removed Area: Remove Sibling Blocks Behind Extract Data:

▼ Plugins

Plugin Name	Plugin Parameter
Referer	
UserAgent	
LinkGenerator	
Proxy	
SessionParameter	
ContentFilter	

Encoding: Crawl Deep: Crawl Frequency:

click here, select content area on page



Administration

?? ?

Shutdown The Server

Data

22/07/2009

Previous 1/1 Next

Common

Music

News

Product

Channel

http://www.mainelegislature.org/legis/bills/bills_124th/b

Directory of Papers

LD 400 - 599

[LD 400. SP0142](#)
An Act To Implement the Recommendations of the Blue Ribbon Commission To Study Long-term Home-based and Community-based Care

[LD 401. HP0297](#)
An Act Creating a Probationary Period for County Corrections Officials

[LD 402. HP0298](#)
An Act To Maintain and Improve ATV Trail Access

[LD 403. HP0299](#)
An Act To Increase the Minimum Wage

[LD 404. HP0300](#)
An Act To Promote Property Tax Relief for Maine's Senior Citizens

[LD 405. SP0149](#)
An Act To Revitalize the Care of Children for the Children's Adult

- HEAD-0
- BODY-0

select node or select text

BODY[0]

Administration

?

i

<<



Data

22/07/2009

Previous 1/1 Next

- Common
- Music
- News
- Product

Administration

Channel

http://www.mainelegislature.org/legis/bills/bills_124th/b

[LD 400, SP0142](#)
 An Act To Implement the Recommendations of the Blue Ribbon Commission To Study Long-term Home-based and Community-based Care

[LD 401, HP0297](#)
 An Act Creating a Probationary Period for County Corrections Officials

[LD 402, HP0298](#)
 An Act To Maintain and Improve ATV Trail Access

[LD 403, HP0299](#)
 An Act To Increase the Minimum Wage

[LD 404, HP0300](#)
 An Act To Promote Property Tax Relief for Maine's Senior Citizens

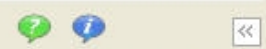
[LD 405, SP0149](#)
 An Act To Reevaluate the Scope of Coverage for the Childless Adult Waiver Program

[LD 406, SP0150](#)
 An Act To Allow a Retailer To Offer To Pay Sales Tax on Items and Services Sold by That Retailer

[LD 407, SP0151](#)
 Resolve To Change the Uses for Funds for Conservation Programs

- HEAD-0
- BODY-0
 - P-0
 - DL-0
 - DT-0
 - DD-0
 - DT-1
 - DD-1
 - DT-2
 - DD-2
 - DT-3
 - DD-3
 - DT-4
 - DD-4
 - DT-5
 - DD-5
 - DT-6
 - DD-6
 - DT-7
 - DD-7
 - DT-8
 - DD-8
 - DT-9
 - DD-9
 - DT-10

BODY[0].DL[0]



thuann:9245 - VietSpider - Information Exploitation System

Channel

22/07/2009

Previous 1/1 Next

Common

Music

News

Product

http://www.mainelegislature.org/legis/bills/bills_124th/b

An Act To Make Minor Substantive Changes to the Tax Laws

- DT-1
- DD-1
- DT-2
- DD-2
- DT-3
- DD-3
- DT-4
- DD-4
- DT-5
- DD-5
- DT-6
- DD-6
- DT-7
- DD-7
- DT-8
- DD-8
- DT-9
- DD-9
- DT-10
- DD-10
- DT-11
- DD-11
- DT-12
- DD-12

BODY[0].DL[0].DD[*] **edit your node**

BODY[0].DL[0].DT[*]

BODY[0].DL[0].DD[1]

add edited node

finish selection node

Administration

thuann:9245 - VietSpider - Information Exploitation System

Data

22/07/2009

Previous 1/1 Next

Common

Music

News

Product

Channel Explorer Category Search

XML

Common

Addresses

Mainelegislature.org

Thomasnet

Zipcodedatabase

Channel Name(*): Mainelegislature.org History Version

Home Page(*): http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1400.asp

Crawled Link Pattern: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/*.asp

Sample Page: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1400.asp

Link Pattern: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/*.asp

Extract Type: ROW Removed Area: Remove Sibling Blocks Behind Extract Data:

Plugins

Plugin Name	Plugin Parameter
Referer	
UserAgent	
LinkGenerator	
Proxy	
SessionParameter	
ContentFilter	

Encoding: ASCII Crawl Deep: 2 Crawl Frequency: 48

Save Verify Reset

Shutdown The Server

each row as a record or the whole site as a record, select here

format xml output, click here

Data

22/07/2009

Previous 1/1 Next

Common

Music

News

Product

Channel

An Act To Designate July 12th as Wyeth Day

DT-0

DD-0

define properties, input name



DD[0]

DEFAULT

Data

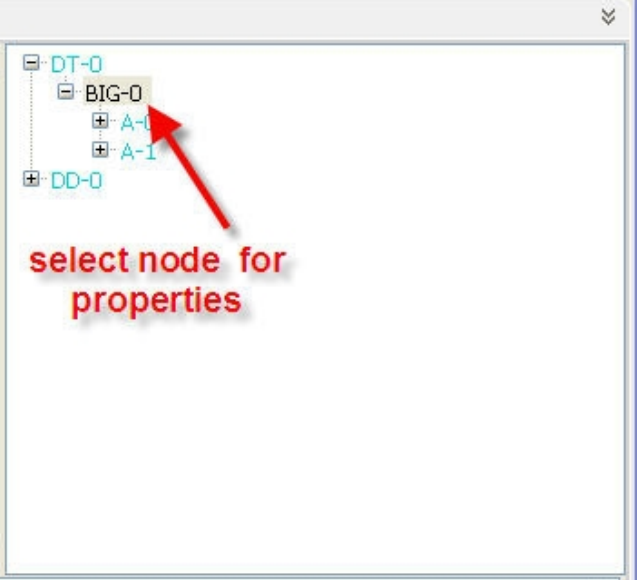
22/07/2009

Previous 1/1 Next

- Common
- Music
- News
- Product

Channel

An Act To Designate July 12th as Wyeth Day



+
name
description

DD[0]

DT[0].BIG[0]

node path of "name" property

thuannd:9245 - VietSpider - Information Exploitation System

Data

22/07/2009

Previous 1/1 Next

Common

Music

News

Product

Administration

Channel

XML

Explorer Category Search

Common

Addresses

Mainelegislature.org

Thomasnet

Zipcodedatabase

Channel Name(*): Mainelegislature.org History Version

Home Page(*): http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1400.asp

Crawled Link Pattern: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/*.asp

Sample Page: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1400.asp

Link Pattern: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/*.asp

Extract Type: ROW Removed Area: Remove Sibling Blocks Behind Extract Data:

▼ Plugins

Plugin Name

Referer

UserAgent

LinkGenerator

Proxy

SessionParameter

ContentFilter

Plugin Parameter

verify your config and save config

Encoding: ASCII Crawl Deep: 2 Crawl Frequency: 48

Save Verify Reset

Shutdown The Server

thuannd:9245 - VietSpider - Information Exploitation System

Data

Administration

- Channel
- Crawler**
- Server

Channel

XML

Explorer Category Search

Common

- Addresses
- Mainelegislature.org**
- Thomasnet
- Zipcodedatabase

Channel Name(*): Mainelegislature.org History Version

Home Page(*): http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1400.asp

Crawled Link Pattern: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/*.asp

Sample Page: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1400.asp

Link Pattern: http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/*.asp

Extract Type: ROW Removed Area: Remove Sibling Blocks Behind Extract Data:

Plugins

Plugin Name	Plugin Parameter
Referer	
UserAgent	
LinkGenerator	
Proxy	
SessionParameter	
ContentFilter	

Encoding: ASCII Crawl Deep: 2 Crawl Frequency: 0

Save Verify Reset

Shutdown The Server

crawl data, click here

thuannd:9245 - VietSpider - Information Exploitation System

Data

Administration

- Channel
- Crawler
- Server

Channel Crawler

Order	Channels	O...	Title	Date
1				

Explorer Search

Group: XML Add all

Category	Source
Common	Addresses
Music	Mainelegislature.org
News	Thomasnet
Product	Zipcode database

select your config, add config to crawl list

Add Sources Close

Start Crawling Crawl Channel Remove Channel

Crawling Threads

add your config to crawl list

Data
Administration



Channel Crawler

Order	Channels	O...	Title	Date
1	Mainelegislature org.XML.Com...			

start crawl data, click here



Start Crawling Crawl Channel Remove Channel

Crawling Threads

Data

Administration



Channel



Crawler



Server

Channel Crawler

Order	Channels	O...	Title	Date
1	Mainelegislature.org.XML.Com...	11	LD 1409, HP0985	09:48:13
		10	LD 1408, HP0984	09:48:13
		9	LD 1407, HP0983	09:48:13
		8	LD 1406, HP0982	09:48:13
		7	LD 1405, SP0508	09:48:13
		6	LD 1404, SP0507	09:48:13
		5	LD 1403, SP0506	09:48:13
		4	LD 1402, HP0981	09:48:13
		3	LD 1401, HP0980	09:48:13
		2	LD 1410, HP0986	09:48:13
		1	LD 1400, HP0979	09:48:12


your data

Stop Crawling

Crawl Channel

Remove Channel

Crawling Threads

Executor 0-XML-Common-Mainelegislature.org-http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents200.asp

Executor 1-XML-Common-Mainelegislature.org-http://www.mainelegislature.org/legis/bills/bills_124th/billtexts/contents1.asp