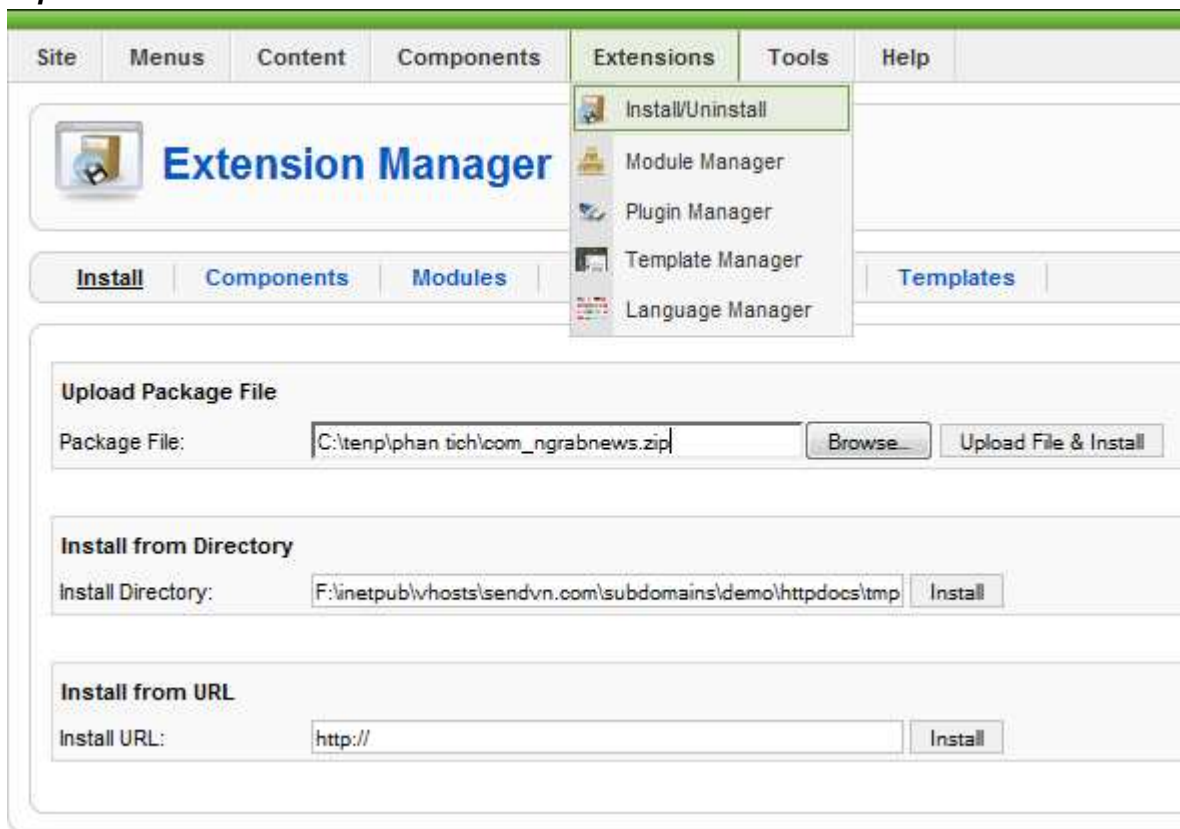


Hướng dẫn cài đặt component lấy tin tự động

1. Cài component:

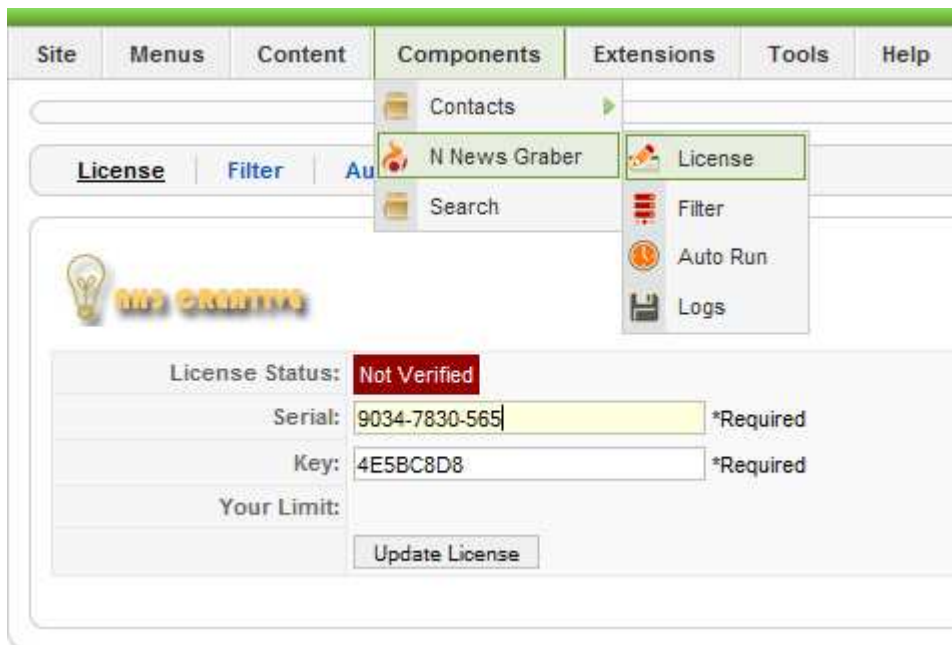
Đăng nhập vào phần quản trị của trang web, chọn menu “**Extension -> Install/Uninstall**”, Chọn file “**com_ngrabnews.zip**” trên máy tính của bạn và nhấn nút “**Upload File & Install**”.



Khi cài đặt xong hệ thống sẽ tự tạo thư mục “**grabnews**” trong “**images/stories/**” nhằm mục đích lưu trữ các file ảnh lấy về được. Bạn hãy kiểm tra thư mục đó có quyền write chưa, hay đã được tạo chưa, nếu chưa có thư mục đó bạn phải tự tạo thủ công.

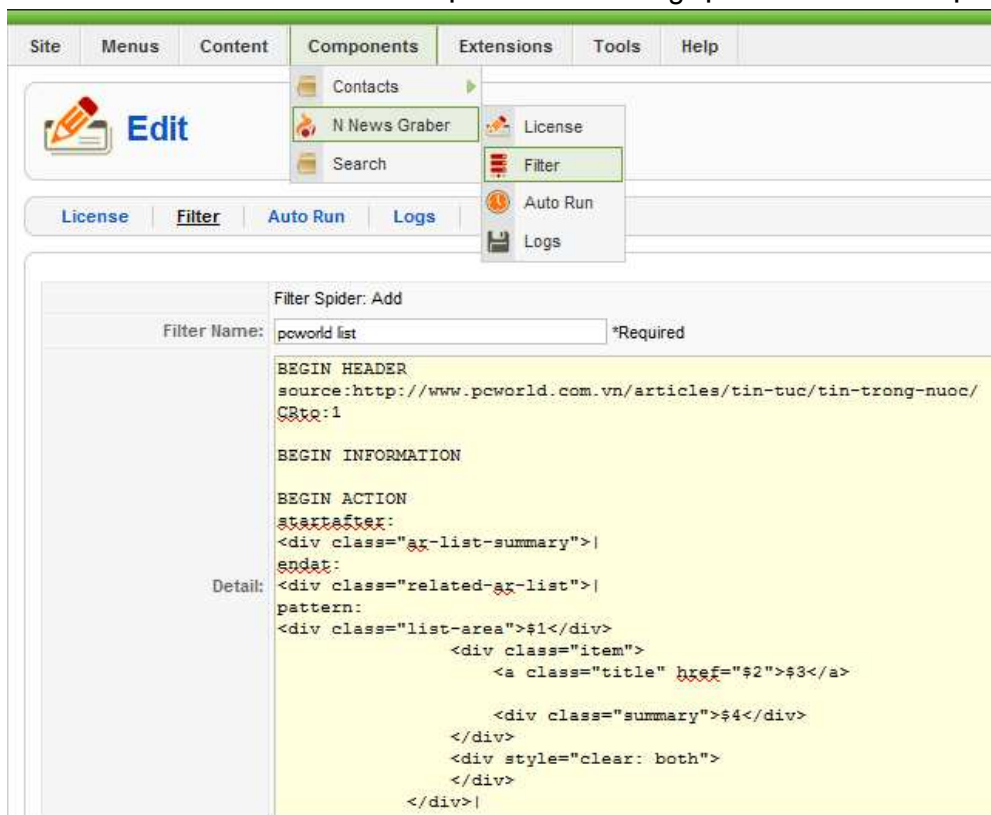
2. Kích hoạt bản quyền:

Chọn menu “**Component -> NNews Grabber -> License**”, điền Serial và Key mà NhutCorp cung cấp cho bạn, và nhấn “**Update License**”. Để cập nhật license thành công, trang web của bạn phải kết nối trực tiếp vào Internet và không bị firewall chặn các kết nối vào ra.








3. Tạo filter:

Chọn menu **"Component -> NNews Grabber -> Filter"**, nhấn nút **"New"** để thêm filter mới. Mở nội dung file Filter mẫu mà chúng tôi gửi bạn copy toàn bộ vào phần Detail, phần Filter Name bạn đặt sao cho dễ phân biệt giữa các Filter, bạn có thể lấy tên file cho dễ nhớ. Nhấn **"Save"** để lưu lại. Thao tác tương tự với file filter còn lại.



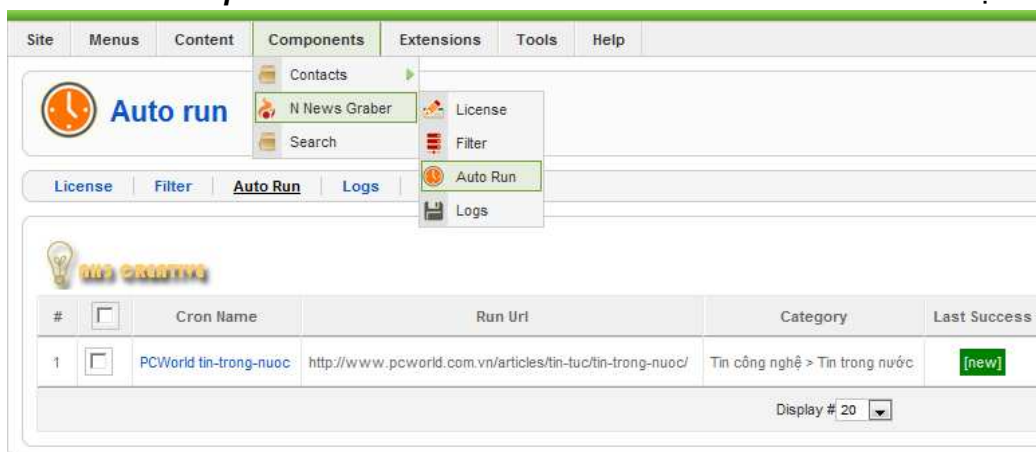
Sau khi save xong bạn có thể nhấn link **“Run”** của mỗi Filter để kiểm tra Filter đó hoạt động tốt không. Bạn phải đảm bảo host của bạn kết nối được trang web đặt trong nội dung filter.

Preview:

SHORT	DETAIL	TITLE
	/articles/tin-tuc/tin-trong-nuoc/2010/08/1220898/vho-vn-dia-chi-truy-van-thong-tin-y-te-mien-phi/	
	/articles/tin-tuc/tin-trong-nuoc/2010/08/1220889/tp-hcm-trien-khai-phan-mem-quan-ly-van-ban-ho-so-cong-viec/	
	/articles/tin-tuc/tin-trong-nuoc/2010/08/1220872/aptech-dua-dien-toan-dam-may-vao-chuong-trinh-dao-tao/	
	/articles/tin-tuc/tin-trong-nuoc/2010/08/1220873/khai-mac-dai-hoi-ix-hoi-nha-bao-viet-nam/	
	/articles/tin-tuc/tin-trong-nuoc/2010/08/1220868/viet-nam-co-5-nha-cung-cap-dich-vu-chu-ky-so/	

4. Thiết lập Cron:

Vào menu **“Component -> NNews Grabber -> Filter”** nhấn nút **“New”** để tạo mới



The screenshot shows the NNews Grabber application window. The 'Components' menu is open, showing 'N News Grabber' and 'Filter'. The 'Filter' sub-menu is also open, showing 'License', 'Filter', 'Auto Run', and 'Logs'. The 'Auto Run' tab is selected, displaying a table of cron jobs.

#	Cron Name	Run Url	Category	Last Success
1	PCWorld tin-trong-nuoc	http://www.pcworld.com.vn/articles/tin-tuc/tin-trong-nuoc/	Tin công nghệ > Tin trong nước	[new]

At the bottom right, there is a 'Display # 20' dropdown menu.

Cài đặt Cron chạy tin <http://www.pcworld.com.vn/articles/tin-tuc/tin-trong-nuoc/> hoặc các phần tin khác giống cấu trúc như vậy, cứ 4h chạy 1 lần.

Auto Run Manager: Add

Publish: ☐ No ☒ Yes

Cron Name: PCWorld tin-trong-nuoc *Required

Group Check: [No Group] ▼

Filter: pworld list ▼ *Required

Run Url: w.poworld.com.vn/articles/tin-tuc/tin-trong-nuoc/ *Required

Section: Tin công nghệ ▼ *Required

Category: Tin trong nước ▼ *Required

f_TITLE: TITLE *Required

f_INTROTEXT: SHORT *Required

f_FULLTEXT: DETAIL *Required

Detail Filter: pworld detail ▼ *Required

f_UNIQUE: DETAIL *Required

Image Intro

Get Detail Image: ☒ No ☐ Yes

Thumb Width: 120

Thumb Height: 120

Image Align: left

Horizontal Spacing: 3

Vertical Spacing: 3

Image Border: 0

Image Detail

Detail Width: 650

Detail Height: 650

Image Align:

Horizontal Spacing:

Vertical Spacing:

Image Border:

Article Setting

Created By: admin ▼ *Required

Show intro: ☒ No ☐ Yes

Auto Publish: ☐ No ☒ Yes

Auto FrontPage: First Item ▼

Source From: Nguồn: PCWorld VN

Auto Fix HTML: ☐ No ☒ Yes

Remove Style: ☒ No ☐ Yes

Remove Internal Link: ☐ No ☒ Yes

Tag Allow:

<table><tbody><thead><tfoot><tr><th><td><colgroup>

<col><p>
<hr><blockquote><i><u><sub><sup><st

rong><tt><var><code><xmp><cite><pre><abbr><acr

onym><address><samp><fieldset><legend><a><h1>

<h2><h3><h4><h5><h6><dl><dt><frame

<button><textarea><div>

* Add your new tag and default tag (preview detail filter to found new tag) or leave blank to use default tag:

<table><tbody><thead><tfoot><tr><th><td><colgroup>

<col><p>
<hr><blockquote><i><u><sub><sup><st

rong><tt><var><code><xmp><cite><pre><abbr><acr

onym><address><samp><fieldset><legend><a><h1>

<h2><h3><h4><h5><h6><dl><dt><frame

<button><textarea>

Auto Keyword: ☐ No ☒ Yes

Black Keyword:

* Auto remove this form keyword, exp: hack,hot girls,sex

Auto Run Setting

Minute(s):

Every Minute

Every Other Minute

Every Five Minutes

Every Ten Minutes

Every Fifteen Minutes

0

1

2

3

4

Hour(s):

Every Hour

Every Other Hour

Every Four Hours

Every Six Hours

0 = 12 AM/Midnight

Months(s):

Every Month

January

February

March

April

Run at:

Day(s):

Every Day

1

2

3

4

Weekday(s):

Every Weekday

Sunday

Monday

Tuesday

Wednesday

UNIX CronTab: 2 */4 * * * * *Required

Last Run:

Updated:

Created:


4






Giải thích một số thông số chính:

- **Group Check:** Kiểm tra tin đã lấy rồi theo nhóm, ngoài kiểm tra đơn lẻ của từng Cron, hệ thống còn hỗ trợ theo nhóm để tránh lấy tin trùng với các Cron khác.
- **Filter:** Chọn bản phân tích dạng list, dùng để lấy tiêu đề, hình ảnh, tin ngắn. Bạn có thể nhấn vào biểu tượng tam giác ngược màu xanh bên cạnh để xem thử.
- **Run Url:** Liên kết cần lấy.
- **Section, Category:** chọn khu vực lưu tin lấy về.
- **Detail Filter:** Bảng phân tích chi tiết trang tin, dùng để lấy thông tin đầy đủ.
- **Image Intro:** Tự động xử lý ảnh cho tin ngắn, bạn có thể chọn "**Get Detail Image**" trong trường hợp tin ngắn ở nguồn tin không có hình, chức năng này lấy 1 ảnh đầu tiên trong tin chi tiết đưa ra ngoài tin ngắn, trang tin bạn sẽ đẹp hơn nếu có hình ảnh đầy đủ.
- **Image Detail:** Xử lý ảnh trong nội dung chi tiết.
- **Auto Fix HTML:** Tự động định dạng lại thẻ HTML trong nội dung lấy về.
- **Remove Style:** Tự động gỡ bỏ các style của site nguồn, nội dung sẽ hiển thị chuẩn với style của site bạn.
- **Remove Internal Link:** Tự động gỡ bỏ các link liên quan đến site nguồn.
- **Tag Allow:** Gỡ bỏ các tag HTML không cần thiết chỉ sử dụng các tag cho phép, ngoài các tag mặc định nếu bạn muốn cho phép thêm tag nào đó thì copy các tag mặc định vào rồi thêm tag của bạn.
- **Auto Keyword:** tự động trích xuất từ khóa và mô tả trang web trên site nguồn vào nội dung bài viết giúp SEO tốt hơn.
- **Black Keyword:** Gỡ bỏ các từ không mong muốn.
- **Auto Run Setting:** Thiết lập giờ tự động kiểm tra và lấy tin về, tùy mức độ cập nhật của site nguồn mà bạn điều chỉnh cho phù hợp, tránh thiết lập thời gian ngắn quá làm nghẽn hệ thống của bạn.


5. Hướng dẫn chạy Cron:


Trước khi chạy Cron bạn có thể test thử bằng nút "**Run Now**", nhấn vào lần đầu hệ thống sẽ lấy tin ngắn, nhấn vào lần 2 hệ thống sẽ lấy thông tin chi tiết. Xin lưu ý rằng, do tính năng test chạy nhiều tin cùng lúc nên có thể gặp time out do qui định thời gian thực thi trên host của bạn, nhưng bạn cứ an tâm chương trình chỉ đưa tin lấy xong vào hệ thống, bạn chỉ cần chạy lại một lần nữa là được. Bài viết chỉ được publish khi đã lấy đầy đủ bản tin.

 **Auto run**

License | Filter | **Auto Run** | Logs

 **NEW CREATIVE**

#	<input type="checkbox"/>	Cron Name	Run Url	Category	Last Success	UNIX Crontab	Last Run	Count	Publish	Delete
1	<input checked="" type="checkbox"/>	PCWorld tin-trong-nuoc	http://www.pcworld.com.vn/articles/tin-tuc/tin-trong-nuoc/	Tin công nghệ > Tin trong nước	Pending 6 Item(s)	2 */4 *	2010-08-12 16:29:46	View(6) Clear		Delete

Display # 20

Thiết lập Cron trên Plesk, cPanel, Webmin hoặc Linux Shell (chạy mỗi 15 phút):

- Minutes: */15
- Hours: *
- Day: *
- Month: *
- Day of the week: 0
- Command:
**[Duong_dan_thu_muc_web_cua_ban]/index.php?option=com_ngrabnews
&view=cron&task=autorun >/dev/null 2>&1**
(thay phần in đậm bằng đường dẫn phù hợp với bạn)

Thiết lập Cron trên Windows:

- Tải chương trình Wget (Open Source) dùng để đọc URL tại địa chỉ sau <http://www.gnu.org/software/wget/> , đặt file wget.exe trong thư mục "C:\wget".
- Dùng cmd chạy lệnh sau để tạo Schedule trong window:
**schtasks /create /sc minute /mo 15 /tn "NGrabNews [Web_cua_ban]" /RU
"NT AUTHORITY\SYSTEM" /tr "C:\wget\wget.exe -q -O nul http://www.
[Web_cua_ban].com/index.php?option=com_ngrabnews&view=cron&task=a
utorun"**
(thay phần in đậm bằng trang web của với bạn).

*Cám ơn đã đọc tài liệu, mọi thắc mắc xin vui lòng add nick
nhutcorp@yahoo.com để được hướng dẫn thêm.*

Để được hỗ trợ tốt bạn nên đăng ký gói có phí <http://www.web-design-123.com/gioi-thieu/tri-ch-xua-t-thong-tin-tu-do-ng>

*NhutCorp
0907.902.100
www.nhutcorp.com*