**LUCENE IN ACTION**

# CHAPTER 9: Lucene ports

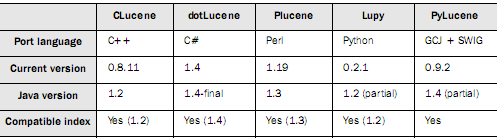
Các nội dung chính được đưa ra và giải quyết trong chương này:

* Sử dụng Lucene ports đến các ngôn ngữ khác nhau.
* So sánh các APIs của các ports , nét đặc trưng, và thực thi.

Trong những năm gần đây số người quan tâm và tham gia vào các dự án liên quan đến Lucene ngày càng đông . Lucene được đưa vào trong thực tế dưới dạng thư viện mã nguồn mở Java IR , Java hiện tại là một trong những ngôn ngữ phổ biến trong giới lập trình viên , tuy vậy nhưng không phải ai cũng biết sử dụng chuyên nghiệp Java. May thay , một số port của Lucene được tạo ra chấp nhận cho các ngôn ngữ khác , trong chương này sẽ nói vắn tắt một số port đã và đang được sử dụng .

## Ports’ relation to Lucene

Các port liên quan đến Lucene được thể hiện rõ qua bảng chi tiết sau:



Đặc trưng của mỗi port thì nó là một dự án độc lập với nhau. Với từng port đều có cung cấp hỗ trợ cho những người quan tâm về nó rất kỹ lưỡng ví dụ website , link download port , địa chỉ mail liên hệ , và nhiều thứ khác . Mỗi port còn có giới thiệu nhóm tác giả và những lập trình viên đang phát triển nó.

Hầu hết các port đều đang cố gắng đồng bộ hóa mã nguồn với những phiên bản mới nhất của Lucene.

## Clucene

Clucene là mã nguồn mở dùng để phát triển cho port của Apache Lucene từ ngôn ngữ C++. Nó được phát hành với giấy phép LGPL và địa chỉ quan tâm <http://sourceforge.net/projects/clucene> . Tại thời điểm viết sách phiên bản hiện tại là 0.8.11 dựa trên phiên bản 1.2 của Lucene. Tới thời điểm hiện tại thì đã cho ra phiên bản 0.9.21b , ngày cập nhật mới nhất là ngày 23/10/2008 . Các gói của CLucene được tích hợp vào cũng giống như trong phiên bản Lucene , gồm cả các gói để chấp nhận cho CLucene được sử dụng với các ngôn ngữ khác. Hiện tại có các gói cho PHP, .Net , Dynamic Link Library(DLL) được chia sẻ giữa các ngôn ngữ với nhau.

* Supported platforms: CLucene được khởi tạo từ Microsoft Visual Studio nhưng giờ nó còn được compiles trong GCC, MinGW32, Borland C++ compiles. Ngoài hệ điều hành Windows , CLucene còn được build thành công trong hệ điều hành Red Hat 9, Mac OS X . Các nhà phát triển còn đang thực hiện hỗ trợ thêm với kiến trúc ADM64, FreeBSD.
* API compatibility: CLucene API tương tự như Lucene .Vấn đề còn đang được giải quyết đó là những ký tự ngoài dãy ASCII .
* Unicode support : được viết dựa trên nền Linux cũ chưa hỗ trợ cho Unicode , vì vậy CLucene không hỗ trợ tốt với các ký tự Unicode . Trong tương lai gần các phiên bản mới có thể giải quyết vấn đề tùy ý lựa chọn thêm vào các thư viện Unicode.
* Users: trên thực tế các nhóm phát triển CLucene thì nhỏ và các bước thực hiện dự án gặp nhiều khó khăn đặc trưng khi thực hiện. Awasu là một người quản lý có kiến thức về các công cụ sử dụng CLucene (http://www.awasu.com/).

## DotLucene

Lucene.Net được phát hành với giấy phép Apache Software License (ASL) , source code có thể tìm thấy từ link <http://www.sourceforge.net/projects/dotlucene/> , sự phân bổ các gói của dotLucene giống như Lucene. Bao gồm source code , test , một vài ví dụ demo.

Ngoài port dotLucene còn có một port khác của Lucene trên nền tảng .Net là Nlucene , link <http://www.sourceforge.net/projects/nlucene/> phiên bản mới nhất đã là từ năm 2002 và nó không chuyển đổi đầy đủ.

* API compatibility: Dù được viết bằng C# nhưng nó xuất ra mọt API gần giống y hệt Lucene.Vì vậy code được viết cho Lucene có thể được port qua C# với sự tác động nhỏ.Sự khác biệt đó là giới hạn của Java và C# về phong cách đặt tên. Những nơi tên của phương thức trong Java thì viết bằng ký tự viết thường, còn bên .Net thì nó được viết in hoa.
* Index compatibility: dotLucene thì tương đương với Lucene trong index level , một index được tạo ra bởi Lucene có thể đọc được bởi dotLucene và ngược lại.Tuy nhiên do Lucene thường xuyên được cập nhật phiên bản mới cho nên nó chỉ tương tích với nhau trong giới hạn từ phiên bản 1.4 , phiên bản hiện tại của Lucene là 2.4.1 (9/3/2009).
* Users: những user nổi tiếng nhất được tìm thấy trong link

<http://www.lookoutsoft.com/Lookout/> , ví dụ như một người thích thú với Lucene.Net như Beagle <http://www.gnome.org/projects/beagle/> đã xây dựng một GNOME componet cho indexing và searching cho tất cả các loại file bao gồm cả hình ảnh.

## Plucene

Plucene là một port của Lucene được viết trên ngôn ngữ Perl, có thể tìm thấy trong CPAN <http://search.cpan.org/dist/Plucene/> . Phiên bản 1.19 được phát hành năm 2004 , vì vậy nó có sự mới lạ với phiên bản 1.3 của Lucene. Hầu hết công việc đã được làm bởi Simon Cozens.

* API compatibility: Nó là port trực tiếp của Lucene , Plucene lĩnh hội được phần lớn các hàm API , và hiển nhiên sẽ có sự khác nhau về cách thức đặt tên trong code , cấu trúc , modules, classes, method.
* Index compatibility: Theo tác giả của Plucene , index được tạo bởi Lucene 1.3 , và Plucene 1.19 thì tương hợp , có thể đọc được từ nội dung được index bởi lẫn nhau.
* Performance: phiên bản 1.19 thì chậm một cách đáng kể so với Lucene ở phiên bản Java.
* Users: theo một cố vấn của Plucene , Plucene thì được sử dụng bởi Gizmodo (<http://www.gizmodo.com/> ), một site xem xét về các thiết bị điện tử cho khách hàng , hay còn được Twingle sử dụng <http://www.twingle.com> .

## Lupy

Lupy là một port của Lucene 1.2 từ ngôn ngữ Python .Những nhà phát triển chính của Lupy là Amir Bakhtiar và Allen Short . Một vài hàm trong Lucene thì bị lỗi trong Lupy như QueryParser , một vài phân tích , index mergin , locking, và một vài items nhỏ . Dù Lupy là một port của một phiên bản Lucene cũ nhưng nó đang dần hoàn thiện để đưa ra một phiên bản mới gần nhất với phiên bản 1.4 . Phiên bản hiện tại của Lupy là 0.2.1, có thể tìm thấy trên trang <http://www.divmod.org/Home/Projects/Lupy/>.

* API compatibility: Lupy API thì tương đồng với Lucene, có sự giống nhau về các classes, methods. Tuy nhiên chúng ta có thể tạo một IndexWriter với chỉ định một analyzer làm một vài thứ mà chúng ta không thể làm trong Lucene.
* Index compatibility: một index được tạo với Lupy thì tương thích với Lucene. Tính tương hợp thì bị giới hạn bởi một phần của phiên bản , index thì chỉ tương đồng với Lucene phiên bản 1.2 .
* Performance: Giống như Plucene , Lupy là một port trực tiếp từ Lucene gốc, chưa có bí quyết nào ghi rõ trong Lupy thì chắc chắn tối ưu trong thực thi hơn Python port.
* Users: người sử dụng chính của Lupy là Divmod <http://www.divmod.com/> .

## PyLucene

PyLucene là một port gần đây nhất của Lucene. Nó được đăng ký bởi MIT license, và người đứng tên là Andi Vajda người có công đóng góp Berkeley DbDirectory trong code chuẩn của Lucene. Nó được indexing và searching component của Chadler , và có thể mở rộng trong cộng đồng mã nguồn mở PIM , nhưng nó trở thành một project riêng biệt trong năm 2004 . Bạn có thể tìm thấy PyLucene tại link <http://pylucene.osafoundation.org/> .

* API compatibility: Không phải tất cả các hàm của Lucene thì dùng được trong PyLucene. Từ khi PyLucene trở thành một project độc lập , người sử dụng có nhiều yêu cầu hơn từ nó, Adi và nhóm phát triển từ từ trình bày nhiều hơn về Lucene API thông qua SWIG . Các cấu trúc liên quan các hàm API thì hầu như giống với Lucene , dễ dàng cho người sử dụng Lucene học cách sử dụng PyLucene .
* Index compatibility: Vì sự tự nhiên của PyLucene (“compiler and SWIG gymnastics”) , index của nó tương thích với Lucene.
* Performance: Mục đích của PyLucene không thể là port nhanh nhất của Lucene nhưng có thể là port gần giống nhất.
* Users: do nó là port mới của Lucene nên chưa có một public user , tuy nhiên có một project được biết đến được xây dựng bởi Chandle <http://www.osafoundation.org/> .