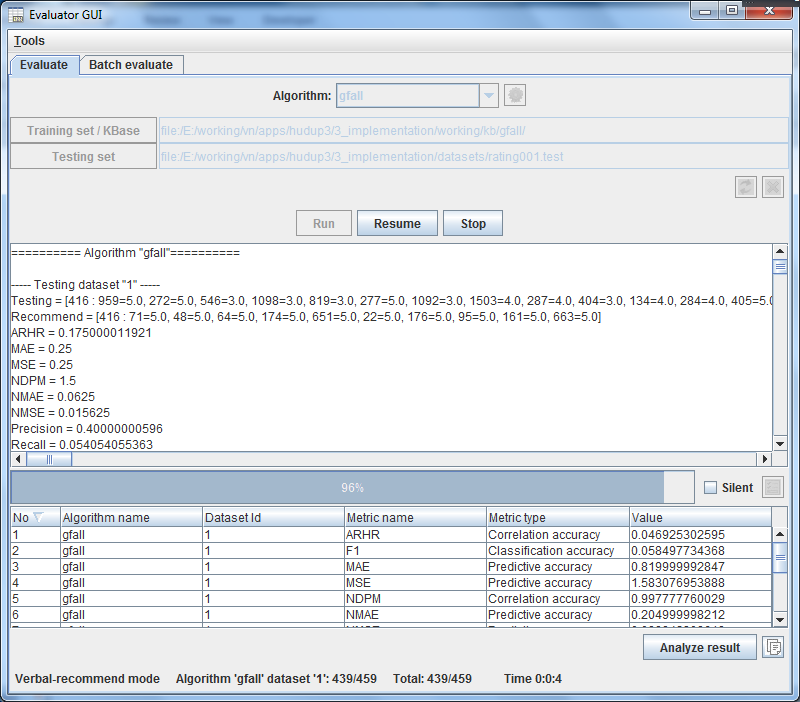
*Bản mô tả sản phẩm*

**Nền tảng hỗ trợ phát triển giải pháp khuyến nghị trong thương mại điện tử**

**Nhà phát triển sản phẩm: Công ty TNHH MTV Lập trình Hướng Dương**

**Phiên bản: 11.0**



**Bản mô tả**

Sáng chế là “*nền tảng hỗ trợ phát triển giải pháp khuyến nghị trong thương mại điện tử*”. Đây là một phần mềm máy tính, một kiến trúc khung, một “hệ điều hành” trong lĩnh vực thương mại điện tử nhằm giúp các nhà khoa học, các nhà phát triển phần mềm xây dựng các giải pháp khuyến nghị dựa trên nền tảng này. Thuật ngữ khuyến nghị (hay tư vấn) đề cập đến các giải pháp (hay thuật toán) công nghệ thông tin có mục đích giới thiệu các sản phẩm mà người mua có thể yêu thích trong một trang web thương mại điện tử nhằm tăng doanh số bán hàng.

Bạn cần phát triển một giải pháp giới thiệu các mặt hàng mà người dùng có thể yêu thích trong một trang web thương mại điện tử. Bạn là một nhà khoa học, sau nhiều năm nghiên cứu, bừng dậy một thuật toán khuyến nghị mới. Ý tưởng của bạn hay, giải pháp của bạn rất hữu ích, bạn hết sức hào hứng nhưng:

1. Bạn đối mặt với những tính toán phức tạp trong xử lý dữ liệu, những mô hình đa dạng trong lĩnh vực khuyến nghị. Thay vì bắt tay ngay vào việc hiện thực hóa ý tưởng, bạn tốn rất nhiều thời gian và công sức ít nhất từ 1 – 2 năm để giải quyết những phức tạp này.
2. Bạn không thể đánh giá được chất lượng giải pháp của bạn theo những tiêu chuẩn phổ biến.
3. Bạn không có môi trường mô phỏng ứng dụng thương mại điện tử để kiểm định tính khả thi của giải pháp bạn đưa ra.

Sáng chế “*nền tảng hỗ trợ phát triển giải pháp khuyến nghị trong thương mại điện tử*” giúp bạn giải quyết 3 vấn đề trên một cách hoàn mỹ và bạn sẽ:

1. Hiện thức hóa được giải pháp rất nhanh và dễ
2. Kiểm định được giải pháp theo các tiêu chuẩn chất lượng
3. Xác định được tính khả thi của giải pháp trong ứng dụng thực tế

Nền tảng còn có một tính năng ưu việt khác, đó là cung cấp hai thuật toán khuyến nghị tối ưu giúp bạn đối sánh các giải pháp với nhau. Sáng chế nhằm phục vụ bạn – là nhà khoa học hay nhà phát triển phần mềm nhằm giải quyết ba vấn đề cốt lõi trên. Sáng chế xây dựng ba nấc giải pháp:

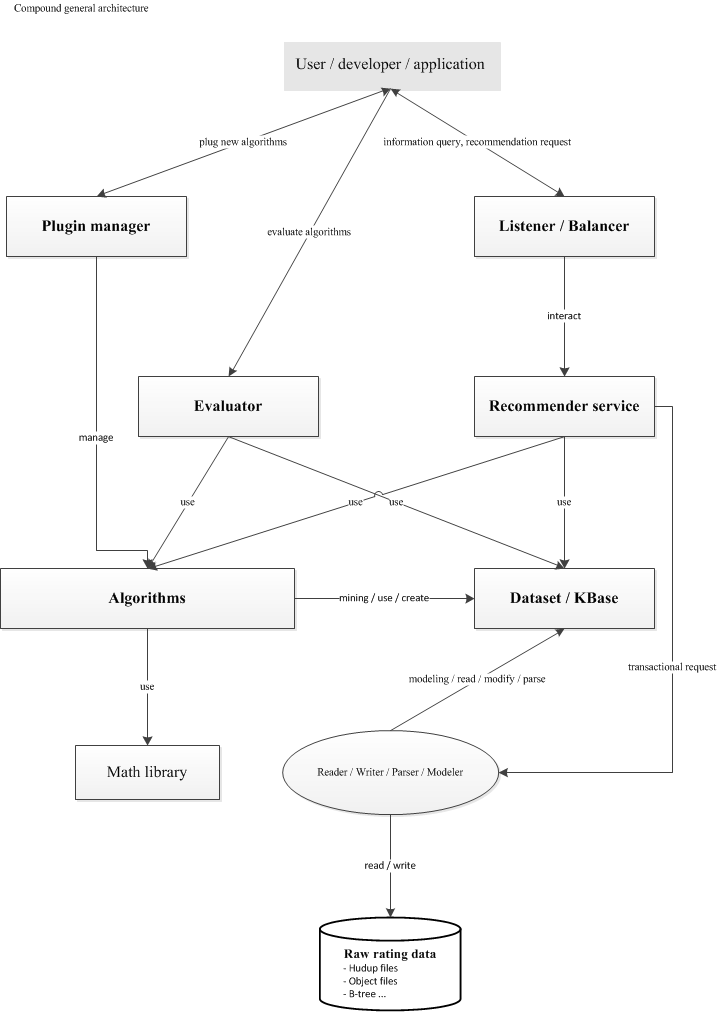
1. Nấc cơ sở: Xây dựng mô hình giải thuật và mô hình dữ liệu giúp bạn sử dụng để tạo ra một sản phẩm phần mềm mới với chi phí thấp nhất.
2. Nấc tiêu chuẩn đo lường: Xây dựng các tiêu chuẩn đánh giá và bộ đánh giá giải thuật giúp bạn xác định được chất lượng giải pháp khuyến nghị.
3. Nấc môi trường mô phỏng: Xây dựng trình chủ khuyến nghị hay môi trường mô phỏng khuyến nghị giúp bạn kiểm định tính khả thi của giải pháp khuyến nghị.

Hiện tại trên thế giới có một số sản phẩm mã nguồn mở hoặc có phiên bản mở có chức năng tương đương với sáng chế bao gồm:

1. *Carleton* (http://www.cs.carleton.edu/cs\_comps/0607/recommend/recommender): được phát triển bởi đại học Carleton, Minnesota, Hoa Kỳ. Phần mềm đơn giản chỉ cài đặt một số thuật toán khuyến nghị và đánh giá thuật toán trên độ đo RMSE. Phần mềm mang tính ví dụ chưa trở thành một nền tảng hay một hệ thống khuyến nghị. Tuy nhiên điểm độc đáo của Carleton là áp dụng trong học tập, giới thiệu những khóa học cho sinh viên dựa trên học bạ của họ. Cấu trúc lớp lập trình trong Carleton tương đối rõ ràng.
2. *Cofi* (http://www.nongnu.org/cofi/): hiện đã ngưng hỗ trợ. Điểm độc đáo của phần mềm này viết trên nền Java nên chạy đa nền đa hệ. Phần mềm đơn giản chỉ cài đặt và đánh giá một số thuật toán, chưa trở thành một nền tảng hay một hệ thống khuyến nghị.
3. *Colfi* (http://colfi.wz.cz/): được phát triển bởi giáo sư Lukáš Brožovský, đại học Charles, Prague. Phần mềm xây dựng một trình chủ khuyến nghị cho dịch vụ hẹn hò. Quy mô của phần mềm này lớn hơn Carleton và Cofi. Phần mềm cài đặt và đánh giá một số thuật toán lọc cộng tác và xây dựng một trình chủ khuyến nghị. Tuy nhiên chưa có tùy biến cho giải thuật và các độ đo, trình chủ còn khá sơ sài. Phần mềm mang nhiều tính nghiên cứu, chưa thể trở thành một nền tảng hay hệ thống khuyến nghị vững chắc. Tuy nhiên điểm độc đáo của phần mềm thể hiện được dịch vụ hẹn hò qua tương tác trang web – máy chủ.
4. *Crab* (http://muricoca.github.io/crab/): trình chủ khuyến nghị được viết bằng ngôn ngữ Python, phát triển bởi phòng thí nghiệp Muricoca http://www.muricoca.com/. Ưu điểm của Crab là xây dựng một trình chủ khuyến nghị cùng guồng máy khuyến nghị có cơ chế đánh giá giải thuật. Nhược điểm của Grab so với sáng chế này là không hỗ trợ nhà phát triển xây dựng một giải thuật mới qua 3 giai đoạn: cài đặt, đánh giá và mô phỏng. Hơn nữa, kiến trúc chưa linh động và chưa chuẩn, giải thuật bên trong trình chủ khuyến nghị chưa đặc sắc chỉ xoay quanh giải thuật SVD và láng giềng gần nhất.
5. *Duine* (http://www.duineframework.org/): được phát triển bởi viện Telematica, Novay đây thực sự là một nền tảng khuyến nghị vững chắc có thể cạnh tranh với sáng chế. Kiến trúc rất linh động và rất tốt. Mặc dù quy mô hơi kém so với sáng chế nhưng Duine có ưu điểm là tập trung vào hiệu năng của guồng máy khuyến nghị. Nhược điểm của Duine so với sáng chế là không hỗ trợ nhà phát triển xây dựng một giải thuật mới qua 3 giai đoạn: cài đặt, đánh giá và mô phỏng. Bộ đánh giá thuật toán của Duine chưa được chuẩn hóa, khả năng tùy biến chưa cao và khá sơ sài.
6. *easyrec* (http://easyrec.org/recommendation-engine) được phát triển bởi tổ chức IntelliJ IDEA (http://www.jetbrains.com/idea/) và Research Studios Austria, Forschungsgesellschaft mbH (http://www.researchstudio.at/home\_en.html). Điểm độc đáo của easyrec là tính tiện dụng, hỗ trợ tư vấn qua mạng, cho phép bạn nhúng guồng máy khuyến nghị vào trang web qua lời gọi hàm đến easyrec. Nhược điểm của easyrec so với sáng chế là không hỗ trợ nhà phát triển xây dựng thuật toán mới. Hơn nữa phân khúc khách hàng của easyrec hoàn toàn khác với chúng tôi.
7. *GraphLab* (http://graphlab.org) là bộ công cụ đa dụng hỗ trợ lọc cộng tác (một lĩnh vực trong khuyến nghị), gom cụm, thị giác máy tính, phân tích đồ thị… được tài trợ bởi Office of Naval Research, Army Research Office, DARPA và Intel. Bộ công cụ này rất lớn và đa dụng, đây là ưu điểm và cũng là nhược điểm. Một số nhà phát triển khi quen dùng bộ công cụ này trong các lĩnh vực khác sẽ có xu hướng tiếp tục sử dụng khi nghiên cứu sang khuyến nghị. Tuy nhiên, riêng đối với lĩnh vực khuyến nghị, bộ công cụ này hỗ trợ rất hạn chế, chỉ cài đặt một số thuật toán lọc cộng tác chưa thể trở thành một nền tảng hay trình chủ khuyến nghị.
8. *LensKit* (http://lenskit.grouplens.org): viết bằng Java chạy đa nền đa hệ, được phát triển bởi nhóm nghiên cứu GroupLens, đại học Minnesota (http://www.grouplens.org/). Điểm mạnh của LensKit là hỗ trợ việc xây dựng và đánh giá giải thuật khuyến nghị rất tốt, cơ chế đánh giá (evaluation) rất tinh xảo. Điểm yếu của LensKit là chưa hỗ trợ trình mô phỏng hay trình chủ khuyến nghị để các nhà phát triển thử nghiệm trong môi trường tương tác khách chủ. Cấu trúc thư viện lớp còn phân mảnh nhưng có điểm hay là tận dụng môi trường phát triển Maven. Nhìn chung, tuy thiếu trình chủ khuyến nghị nhưng LensKit là một nền tảng rất tốt.
9. *Mahout* (http://mahout.apache.org/) được phát triển bởi tổ chức Apache (http://www.apache.org), là bộ công cụ rất mạnh hỗ trợ khai thác dữ liệu và học máy trong đó có cài đặt một số thuật toán khuyến nghị như láng giềng gần nhất. Thử nghiệm giải thuật trên Mahout rất đơn giản, Mahout hướng đến người dùng cuối nhiều hơn nhà phát triển phần mềm. Điểm mạnh và điểm yếu của Mahout rất gần với GraphLab, mục tiêu chủ yếu của Mahout là bộ công cụ đa dụng nên không hỗ trợ hoàn toàn khuyến nghị. Nếu bạn phát triển một phần mềm khai khoáng dữ liệu hay học máy thì nên dùng Mahout, còn bạn muốn tập trung vào khuyến nghị thì nên dùng sáng chế của chúng tôi.
10. *MyMedia* (http:// mymediaproject.codeplex.com): là phần mềm giới thiệu các sản phẩm đa phương tiện như phim, ảnh. Điểm đặc sắc của MyMedia là cài đặt các thuật toán khai thác mạng xã hội, khuyến nghị, cá nhân hóa tập trung vào thông tin giải trí đa phương tiện. MyMedia là một nền tảng khuyến nghị thông tin đa phương tiện rất mạnh hướng đến người dùng cuối là các tổ chức, công ty hoạt động trong lĩnh vực giải trí, đa phương tiện. Điểm yếu của MyMedia là không cài đặt trình chủ khuyến nghị và cơ chế đánh giá giải thuật chưa hiệu quả. Nhìn chung MyMedia hướng đến người dùng cuối. MyMedia viết bằng ngôn ngữ lập trình hiện đại C#, được phát triển bởi tổ chức EU Framework 7 Programme Networked Media initiative cùng với các nhà cộng tác: EMIC, BT, the BBC, Technical University of Eindhoven, University of Hildesheim, Microgenesis and Novay.
11. *MyMediaLite* (http://mymedialite.net) là bộ thư viện nhỏ, gọn cài đặt và đánh giá một số thuật toán khuyến nghị. Điểm mạnh của MyMediaLite là nhỏ gọn nhưng cài đặt khá nhiều các giải thuật khuyến nghị và các độ đo, kiến trúc rất trong sáng. Điểm yếu: chưa cài đặt trình chủ khuyến nghị, các độ đo chưa có khả năng tùy biến, bộ đánh giá giải thuật khá sơ sài.
12. *Recommenderlab* (A Framework for Developing and Testing Recommendation Algorithms – Michael Hahsler 2011) được phát triển bởi Michael Hahsler tài trợ bởi NSF Industry/University Cooperative Research Center for Net-Centric Software & Systems. Recommenderlab là gói mở rộng ngôn ngữ thống kê R với mục tiêu là xây dựng một cơ sở hạ tầng khuyến nghị trên nền R. Điểm đặc sắc của Recommenderlab là tận dụng khả năng xử lý dữ liệu xuất sắc của R. Khả năng đánh giá, so sánh các thuật toán trong Recommenderlab cũng rất tốt. Điểm yếu là chưa phát triển trình chủ khuyến nghị vì Recommenderlab không thể chạy độc lập với R. Recommenderlab phù hợp với nghiên cứu và đánh giá giải thuật trong thời gian ngắn; việc phát triển, đánh giá và thử nghiệm một giải thuật cần một nền tảng mạnh mẽ hơn như MyMedia, Mahout, LensKit, Duine và sáng chế chúng tôi là lựa chọn tốt nhất.
13. *SVDFeature* (http://svdfeature.apexlab.org) viết bằng C++ được phát triển bởi các nhà phát triển Tianqi Chen , Weinan Zhang , Qiuxia Lu , Kailong Chen , Zhao Zheng , Yong Yu. SVD là một giải thuật lọc cộng tác xử lý ma trận rất hiệu quả trong lĩnh vực khuyến nghị. SVDFeature tập trung vào cài đặt giải thuật SVD một cách tốt nhất. SVDFeature không phải là một nền tảng khuyến nghị nhưng có điểm đặc sắc là xử lý dữ liệu ma trận lớn và tăng tóc giải thuật SVD.
14. *Vogoo* (http://sourceforge.net/projects/vogoo/). Sự độc đáo của Vogoo là hướng đến người dùng cuối, cài đặt giải thuật khuyến nghị ngay trên trang web viết bằng ngôn ngữ web PHP, điều này mang đến sự tiện dụng tối đa khi xây dựng trang web thương mại điện tử bằng PHP. Điểm yếu là Vogoo quá sơ sài, không phải là một nền tảng khuyến nghị mặc dù qua khảo sát tôi nhận thấy Vogoo độc đáo nhất, thư viện cực kỳ nhỏ gọn, nếu yêu cầu phát triển nhanh đặt lên hàng đầu, Vogoo là số một.

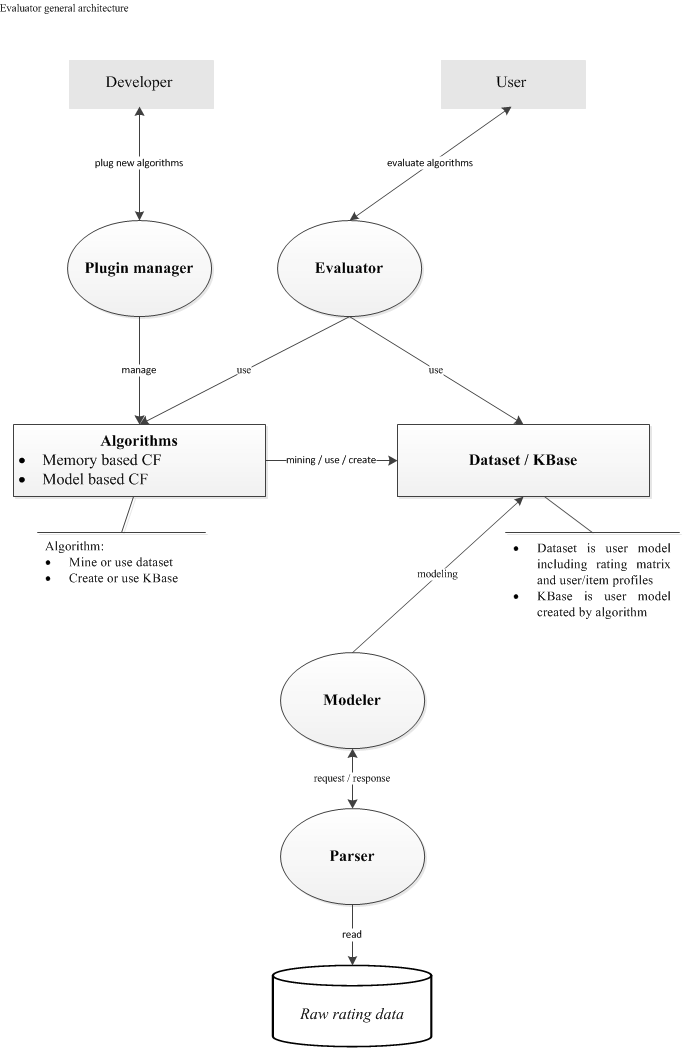
Qua khảo sát 14 phần mềm khuyến nghị tiêu biểu, sáng chế này độc đáo nhất và tối ưu nếu xét theo khía cạnh hỗ trợ xuyên suốt nhà phát triển xây dựng một giải pháp khuyến nghị mới từ khâu cài đặt, khâu kiểm định chất lượng đến khâu thử nghiệm. Hơn nữa kiến trúc sản phẩm tạo thành từ sáng chế rất linh động và tùy biến cao. Các tiêu chuẩn kiểm định (độ đo) và giải thuật đều được chuẩn hóa theo một khuôn mẫu nhưng người dùng vẫn có thể thêm những độ đo và giải thuật mới.

Sáng chế là một phần mềm máy tính được cấu thành từ 3 bộ: bộ quản lý giải thuật và độ đo (plugin manager), bộ đánh giá (evaluator) và dịch vụ khuyến nghị (recommender service) tương ứng với 3 nấc giải pháp mà sáng chế đề xuất. Sau đây là kiến trúc tổng quan của sáng chế:



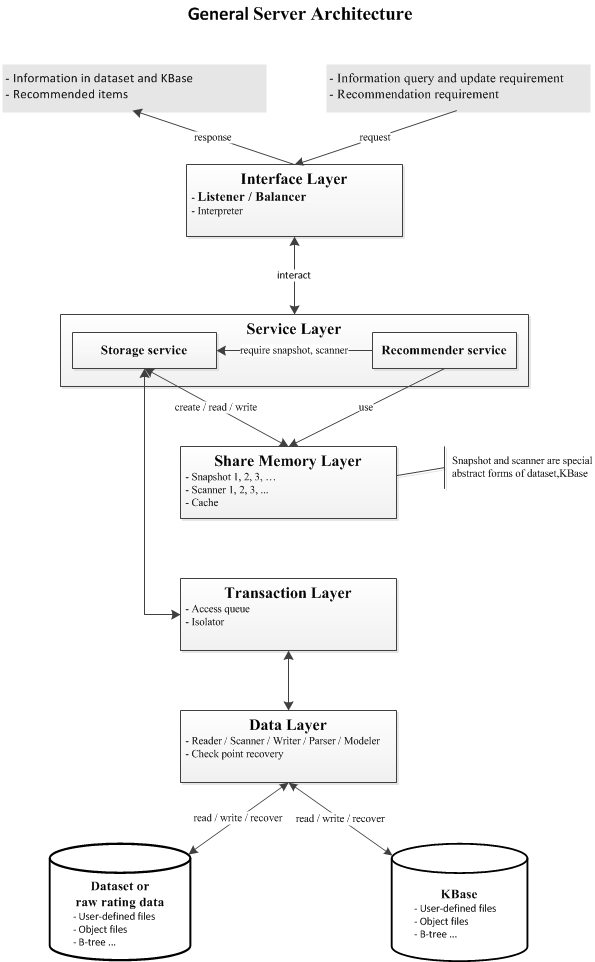
**Hình 1**. Kiến trúc tổng quan

Sau đây là kiến trúc bộ đánh giá (evaluator)



**Hình 2**. Kiến trúc bộ đánh giá (evaluator)

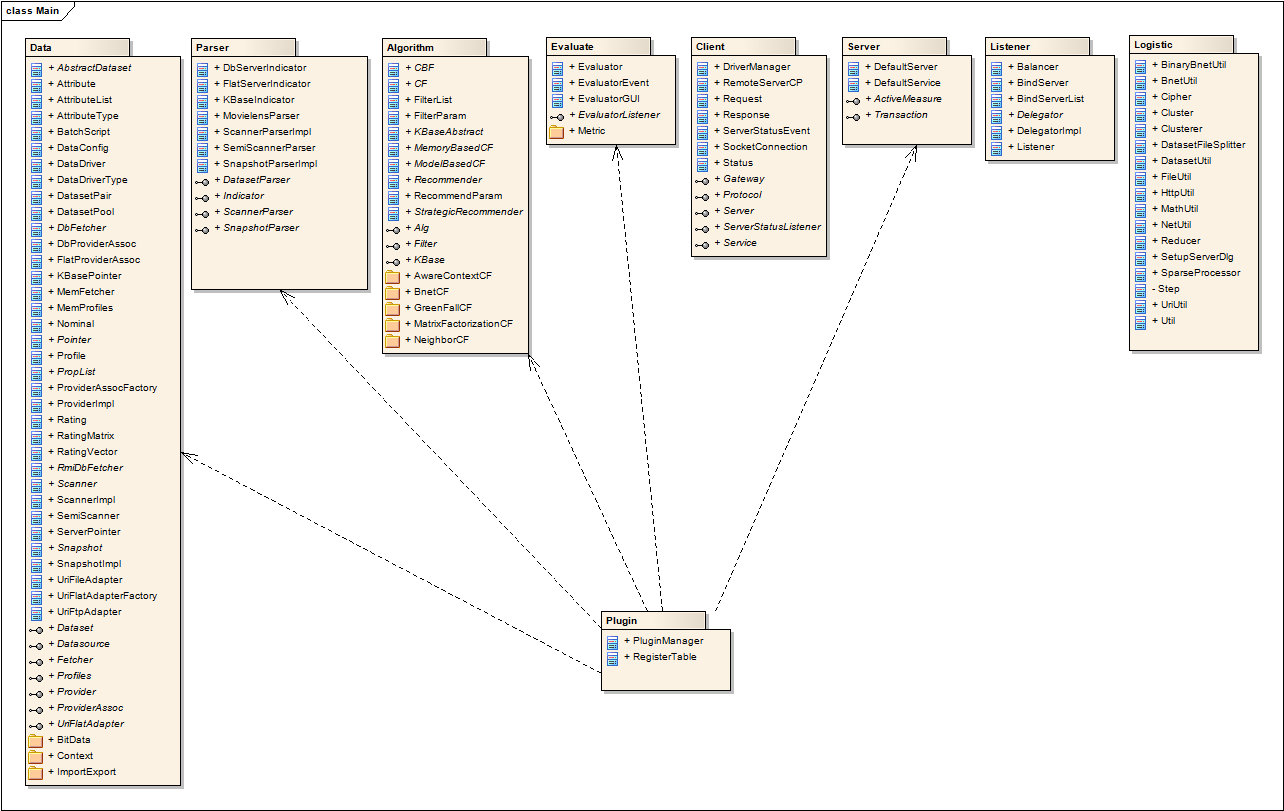
Sau đây là kiến trúc dịch vụ khuyến nghị



**Hình 3**. Kiến trúc dịch vụ khuyến nghị (recommender service or receommender server)

Kiến trúc tổng quan 3 bộ được phân rã thành 9 gói (package):

* *Data* (*dữ liệu*): chuẩn hóa và mô hình hóa dữ liệu ở mức trừu tượng
* *Parser* (*phân tích*): xử lý dữ liệu
* *Algorithm* (*giải thuật*): mô hình giải thuật khuyến nghị
* *Evaluate* (*đánh giá*): đánh giá giải thuật theo các độ đo đã được cài đặt
* *Client* (*khách*), *Server* (chủ) và *Listener* (*bộ nghe*): cung cấp dịch vụ khuyến nghị trong môi trường mạng máy tính khách – chủ
* *Logistic* (*tiện ích*): các tiện ích tính toán và toán học
* *Plugin* (*phích cắm*): quản lý các giải thuật và độ đo



**Hình 4**. Chín gói của sáng chế

Mỗi gói gồm nhiều lớp (class) phần mềm máy tính tạo thành các sơ đồ lớp (kiến trúc lớp) nội tại. Kiến trúc mỗi gói được mô tả trong phụ lục A. Riêng gói *Algorithm* còn cung cấp 2 thuật toán khuyến nghị: lọc cộng tác dựa trên mạng Bayesian (*BnetCF*) và lọc cộng tác dựa trên khai thác tập phổ biến cực đại (*GreenFallCF*). Chi tiết các thuật toán được mô tả trong 2 bài báo phụ lục B.

Sáng chế giúp xây dựng các giải thuật khuyến nghị một cách nhanh chóng và dễ dàng cùng kiểm định tính chính xác của giải thuật qua các độ đo và kiểm định tính khả thi của giải pháp trong môi trường mô phỏng. Giả sử bạn cần cài đặt thuật toán lọc cộng tác Green Fall dựa trên khai thác tập phổ biến cực đại, thay vì viết cả phần mềm lớn tốn rất nhiều thời gian bao gồm xử lý dữ liệu, thi công chi tiết thuật toán, cài đặt chi tiết các độ đo, kiểm định thuật toán, xây dựng môi trường mô phỏng thuật toán trong ứng dụng thời gian thực, bạn chỉ cần sử dụng phần mềm từ sáng chế và thực hiện 3 bước sau:

1. Kế thừa lớp (class) *MemoryBasedCF* hay *ModelBasedCF* trong gói *Algorithm* và cài đặt ý tưởng của bạn trong 2 phương thức (method) *estimate* và *recommend* trong 2 lớp trên.
2. Khởi động bộ đánh giá *Evaluator* (xem hình 1 và 2) để kiểm định và so sánh GreenFall với các thuật toán qua các độ đo đã được cài đặt sẵn. Bạn hoàn toàn có thể thêm vào độ đo mới
3. Cấu hình trình chủ khuyến nghị *Recommender service* (xem hình 1 và 3) để nhúng Green Fall vào trong trình chủ và khởi động trình chủ để kiểm định tính khả thi của Green Fall trong ứng dụng thời gian thực

Các thao tác thực hiện trong 3 bước rất đơn giản, chủ yếu là cấu hình qua giao diện phần mềm, ngoài bước 1 – bạn cần viết những dòng mã nguồn (code line) để thi công ý tưởng. Mô hình thuật toán đã được thiết kế và cài đặt chặt chẽ, tất cả những dòng mã nguồn bạn viết gói gọn trong 2 phương thức *estimate* và *recommend*. Ước lượng thời gian và công sức để cài đặt và kiểm định một thuật toán khuyến nghị khoảng 2 năm, nếu sử dụng sáng chế sẽ rút ngắn khoảng 1 tuần khi ý tưởng của bạn rõ ràng. Điều này có nghĩa chi phí phát triển thuật toán giảm xuống rất nhiều còn khoảng 1% so với tổng chi phí ban đầu.

Bản mô tả sản phẩm gồm 2 phụ lục:

* Phụ lục A: Từ hình 5 đến hình 19 mô tả tất cả các lớp (class) phần mềm thi công kiến trúc này.
* Phụ lục B: Chi tiết 2 thuật toán: lọc cộng tác dựa trên khai thác tập phổ biến cực đại (*Green Fall*) và lọc cộng tác dựa trên suy diễn mạng Bayesian (*Bayesian*).