

PHẦN MỀM NHẬN DIỆN 220 LOÀI CÂY GỖ RỪNG ĐÔNG NAM BỘ, VIỆT NAM

Đặng Lê Anh Tuấn⁽¹⁾, Jérôme Millet⁽²⁾, Nguyễn Lê Xuân Bách⁽³⁾, Trần Ninh,
Pierre Bonnet⁽⁴⁾, Pierre Grard⁽⁴⁾

(1) Trường Đại học Khoa học Tự nhiên, ĐHQG - HCM

(2) Trường Đại học Lyon1, Pháp ; (3) Viện sinh học nhiệt đới (ITB)

(4) Viện nghiên cứu thực vật và sinh tin học về kiến trúc thực vật, Pháp

(Bài nhận ngày 21 tháng 03 năm 2011, hoàn chỉnh sửa chữa ngày 18 tháng 11 năm 2011)

TÓM TẮT : Dựa trên các đặc điểm hình thái dễ nhận ngoài thực địa, đặc biệt là cấu trúc lá và cấu trúc phần mềm IDAO của CIRAD, chúng tôi đã phát triển phần mềm nhận diện cho 220 loài cây gỗ rừng Đông Nam Bộ, Việt Nam, với dự định phiên bản cuối sẽ có 265 loài trong năm 2011. Với giao diện trực quan mô phỏng các đặc điểm hình thái, người sử dụng nhanh chóng đạt được kết quả nhận diện ngay sau mỗi đặc điểm chỉ với thao tác so sánh, chọn đặc điểm phù hợp. Kết quả này cho mỗi loài gồm bản mô tả các đặc điểm, công dụng cũng như các hình ảnh kèm theo. Trong phần kết quả, nếu có nhiều loài có cùng độ tương đồng thì phần mềm sẽ đề nghị các đặc điểm nên thu thập thêm để tiếp tục nhận diện các loài này. Một tiện ích khác của phần mềm là việc giúp người sử dụng nhận biết các lỗi gặp phải trong quá trình nhận diện cũng như đề nghị các phương án chỉnh sửa phù hợp.

Từ khóa: IDAO, cây gỗ, Đông Nam Bộ Việt Nam

GIỚI THIỆU

Các phần mềm nhận diện trực quan, một ứng dụng liên kết phân loại và hình thái học thực vật với các tiện ích của công nghệ thông tin, hiện đang được quan tâm phát triển. Các tác giả đầu tiên phát triển ý tưởng này có thể kể như Goodall [1]; Hall [2]. Năm 1974, Lobanov [3] là người thành công đầu tiên khi đưa ra chương trình Diagnostica-1 chạy trên máy tính Nairi. Năm 1986, Duncan và Meacham [4] viết chương trình MEKA (Multiple Entry Key Algorithm) trên giao diện commandos giúp nhận diện các loài sinh vật với 2 ứng dụng đầu tiên là Angiofam (nhận diện các họ thực vật thuộc nhóm hiển hoa bí từ), Tropifam (nhận

diện các họ thực vật thuộc nhóm hiển hoa bí từ nhiệt đới). Gần đây nhất, hệ thống DELTA [5], được sử dụng nhiều trong việc thành lập các khóa nhận diện sinh vật với sự hỗ trợ của công nghệ thông tin. Tính đến cuối thế kỉ 20, hơn 40 chương trình hỗ trợ nhận diện sinh vật đã được phát triển, trong đó được sử dụng nhiều nhất là Lucid ® và Intkey ® [5]. Được phát triển tại Centre International de Recherche Agronomique pour le Developpement (CIRAD) từ năm 1996, phần mềm IDAO ® (IDentification Assistée par Ordinateur), phần mềm hỗ trợ “tự học” nhận diện đa trạng thái (multimedia self-training species identification software), giúp nhận diện nhanh các loài thực vật dựa trên các hình mô phỏng trực quan với

các mô tả chi tiết, thuật ngữ thực vật học và hình ảnh minh họa kèm theo.

Ở Việt Nam, phát triển phần mềm nhận diện các loài sinh vật là một hướng nghiên cứu mới. Một trong những công trình tâm huyết đầu tiên là “Sinh vật rừng Việt Nam” (SVRVN) của Phùng Mỹ Trung và csv. [10]. Trong phiên bản mới nhất (2.0), phần mềm SVRVN ghi nhận 4500 loài thực vật, 2000 loài động vật và 600 loài côn trùng. Phần kết quả của mỗi loài có 1 ảnh nhỏ kèm theo các mô tả những đặc điểm cơ bản.

Trong bài nghiên cứu này, chúng tôi chọn cấu trúc phần mềm IDAO để phát triển phần mềm nhận diện cho 220 loài cây gỗ rừng Đông Nam Bộ, Việt Nam. Phần mềm IDAO thể hiện nhiều ưu điểm so với các phần mềm nhận diện khác: 1/ nhận diện dựa trên hình ảnh minh họa thay vì các thuật ngữ chuyên môn trừu tượng; 2/ người sử dụng tự do lựa chọn các đặc điểm sẵn có trên mẫu vật chứ không bị gò bó theo trật tự như trong khóa nhận diện; 3/ tính được phần trăm tương thích của các loài còn lại trong kết quả dựa trên các đặc điểm đưa vào; 4/ thể hiện và hỗ trợ người sử dụng chỉnh sửa đặc điểm không phù hợp ở các loài có độ tương thích dưới 100%; 5/ công cụ “Tìm kiếm” giúp người sử dụng chọn đặc điểm tiếp theo; 6/ cho phép phát triển phát triển trên nhiều ngôn ngữ; 7/ trong phần mô tả của mỗi loài, mỗi thuật ngữ chuyên môn được liên kết với định nghĩa và hình minh họa.

PHƯƠNG PHÁP

Thu và xử lý mẫu

Chúng tôi khảo sát, thu mẫu cây thân gỗ tại 3 vườn quốc gia: Bù Gia Mập (Tỉnh Bình Phước), Cát Tiên (Tỉnh Đồng Nai) và Lò Gò Xa Mát (Tỉnh Tây Ninh). Dựa trên bản đồ thăm thực vật, chúng tôi chọn khảo sát các tuyến đại diện cho các kiểu rừng đặc trưng ở mỗi vườn quốc gia, thu mẫu các loài cây gỗ có lá, hoa và trái dọc các tuyến khảo sát. Sử dụng ống nhôm để tìm các loài có hoa, trái, đánh số lên vỏ cây, định tọa độ, ghi nhận nhanh các đặc điểm về sinh cảnh, dạng sống, nhựa mủ, màu hoa. Chúng tôi thu 3-4 mẫu ở mỗi cây bằng kéo cắt cành cao hoặc leo cây thu mẫu. Sau khi được chụp hình, các mẫu vật được ép báo và sấy khô ngay trên thực địa.

Mô tả, định danh mẫu vật, thực hiện tiêu bản thực vật

Tại phòng thí nghiệm, sử dụng kính lúp 2 trường để mô tả chi tiết về các đặc điểm thực vật học của lá, hoa, trái của mỗi mẫu vật. Các đặc điểm thực vật học được sử dụng trong mô tả mẫu vật dựa theo Harris [6]. Dựa trên các mô tả này, chúng tôi xác định tên khoa học của mẫu vật thông qua việc so sánh với các tài liệu phân loại thực vật như Cây cỏ Việt Nam [7], Thực vật chí Đông Dương [8], Thực vật chí Lào, Campuchia và Việt Nam [9].

Lập cơ sở dữ liệu cho phần mềm

Sau khi đã mô tả, định danh, chúng tôi chọn các đặc điểm nhận diện hình thái cơ quan dinh dưỡng và sinh sản, sử dụng chương trình Microsoft Access và phần mềm Corel Draw để lập cơ sở dữ liệu các đặc điểm đã chọn cho phần mềm. Các cơ sở dữ liệu này nhất thiết phải tương thích với nhau. Trong đó, có 2 phần

cơ sở dữ liệu văn bản: a/ bảng thể hiện danh sách loài cùng với các đặc điểm hiện diện (kí hiệu 1) và không hiện diện (kí hiệu 0), b/ bảng mô tả đặc điểm thực vật học và các định nghĩa của các thuật ngữ này và 1 phần dữ liệu hình ảnh (hình chụp của các loài và hình mô phỏng các đặc điểm nhận diện đã chọn). Cả 2 dạng dữ liệu này về sau sẽ được chuyển thành dạng trang mạng liên kết với các thuật ngữ và hình ảnh tương thích.

Thay đổi cấu trúc, giao diện phần mềm IDAO cho tương thích với cơ sở dữ liệu

Khi đã có dữ liệu, chúng tôi thay đổi cấu trúc, và nếu cần thiết là giao diện phần mềm IDAO cho tương thích với cơ sở dữ liệu mới. Việc thay đổi này được thực hiện trên phần mềm Visual Basic.

Phát triển, chạy thử nghiệm, chỉnh sửa và xuất bản phần mềm

Sau khi đã có đầy đủ dữ liệu gồm các trang mạng liên kết với các hình ảnh và định nghĩa tương thích với giao diện và cấu trúc phần mềm, chúng tôi tạo tập tin cài đặt cho phần mềm và chạy thử nghiệm. Trong quá trình chạy thử nghiệm, chúng tôi kiểm tra và chỉnh sửa tính tương thích của giao diện và cấu trúc phần mềm với cơ sở dữ liệu. Bước cuối cùng, sau

khi chạy thử nghiệm đạt yêu cầu, chúng tôi cho xuất bản tập tin cài đặt và các dữ liệu liên quan dưới dạng CDrom.

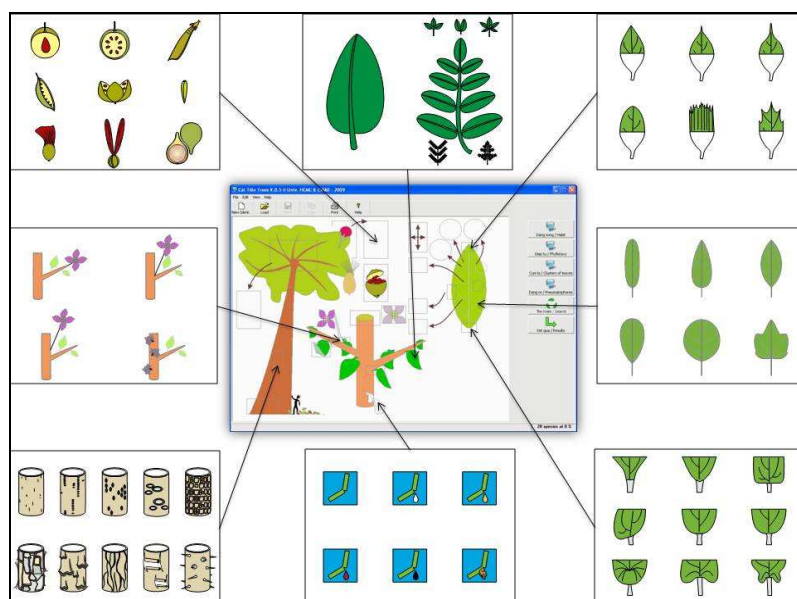
KẾT QUẢ VÀ THẢO LUẬN

Bộ tiêu bản thực vật

Trong tổng số 346 mẫu thu được, hiện chúng tôi đã định danh và làm tiêu bản cho 265 loài thực vật, mỗi loài 3 tiêu bản. Bộ tiêu bản này sẽ được phân bổ và lưu trữ tại thảo tập của Phòng Thực vật, Bộ môn Sinh thái – Sinh học tiến hóa, Khoa Sinh, Trường Đại học Khoa học tự nhiên, Đại học quốc gia Thành phố Hồ Chí Minh và bảo tàng thực vật Paris, Cộng hòa Pháp. Với tọa độ đã được định vị, các loài đã được thu thập có thể được tìm lại hoặc để tiếp tục thu mẫu bổ sung, giúp bộ tiêu bản đầy đủ các bộ phận lá, hoa và trái.

Cơ sở dữ liệu cho phần mềm

Chúng tôi đã phát triển phiên bản Tân Phú, có cơ sở dữ liệu bằng tiếng Pháp và tiếng Việt cho 200 loài và phiên bản cho 20 loài khác ở vườn quốc gia Cát Tiên, tỉnh Đồng Nai. Ở phiên bản Tân Phú, cơ sở dữ liệu bao gồm 185 đặc điểm hình thái thuộc 18 bậc đặc điểm cơ quan dinh dưỡng và 5 bậc đặc điểm cơ quan sinh sản, còn ở phiên bản Cát Tiên là 232 đặc điểm.

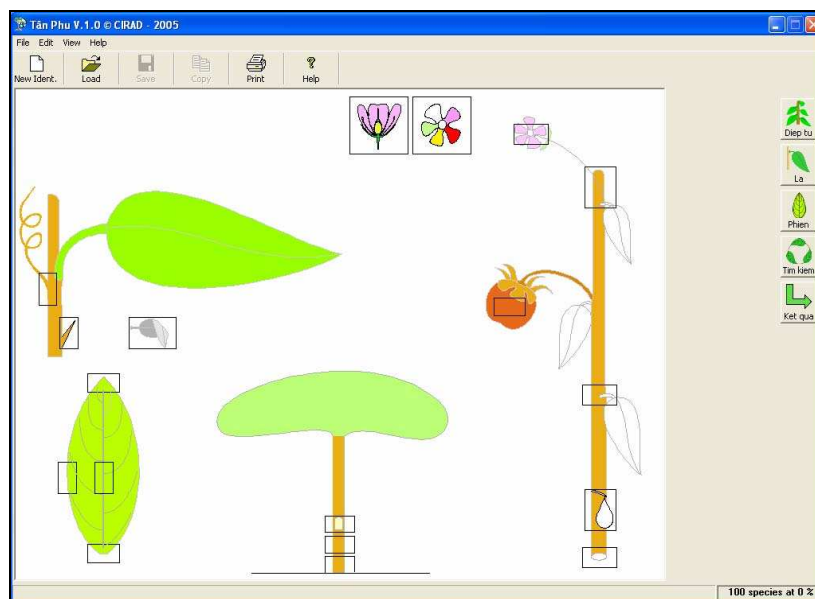


Hình 1. Hình vẽ minh họa một số đặc điểm trong phần mềm Cát Tiên

Ở phiên bản Đông Nam Bộ, chúng tôi chọn giao diện gốc của phiên bản Tân Phú và đã lập cơ sở dữ liệu cho 265 loài, và thêm 40 đặc điểm nhận diện mới ở lá so với phiên bản Tân Phú. Trong phiên bản này, chúng tôi tập trung vào các đặc điểm hình thái lá vì tính sẵn có của cơ quan này. Mẫu lá được sử dụng cho phần mềm có thể thu thập gần như quanh năm hay có thể nhặt trên mặt đất. Theo kết quả nghiên cứu của Đặng Lê Anh Tuấn [11], hệ gân lá có tính ổn định cao, có thể được sử dụng trong quá trình nhận diện mẫu thực vật.

Chúng tôi đã hoàn thành việc xuất bản dưới dạng CDrom cho 200 loài ở phiên bản Tân Phú và chạy thử nghiệm thành công phần mềm nhận diện cho 20 loài ở phiên bản Cát Tiên.

Đối với phiên bản Đông Nam Bộ, chúng tôi hiện đang hoàn tất công đoạn cuối (giao diện phần mềm IDAO cho tương thích với cơ sở dữ liệu và phát triển, chạy thử nghiệm để xuất bản phần mềm). Dự kiến sẽ xuất bản phần mềm nhận diện cây gỗ rừng Đông Nam Bộ, Việt Nam với 265 loài trong tháng 8 năm 2011. Dựa trên danh lục thực vật của một số khu bảo tồn thiên nhiên và vườn quốc gia, ước tính khu vực Đông Nam Bộ có khoảng 800 loài thân gỗ, trong đó 350 loài thân gỗ lớn. Như vậy, phiên bản khi xuất bản hỗ trợ nhận diện được khoảng 30% các loài thân gỗ có trong khu vực.



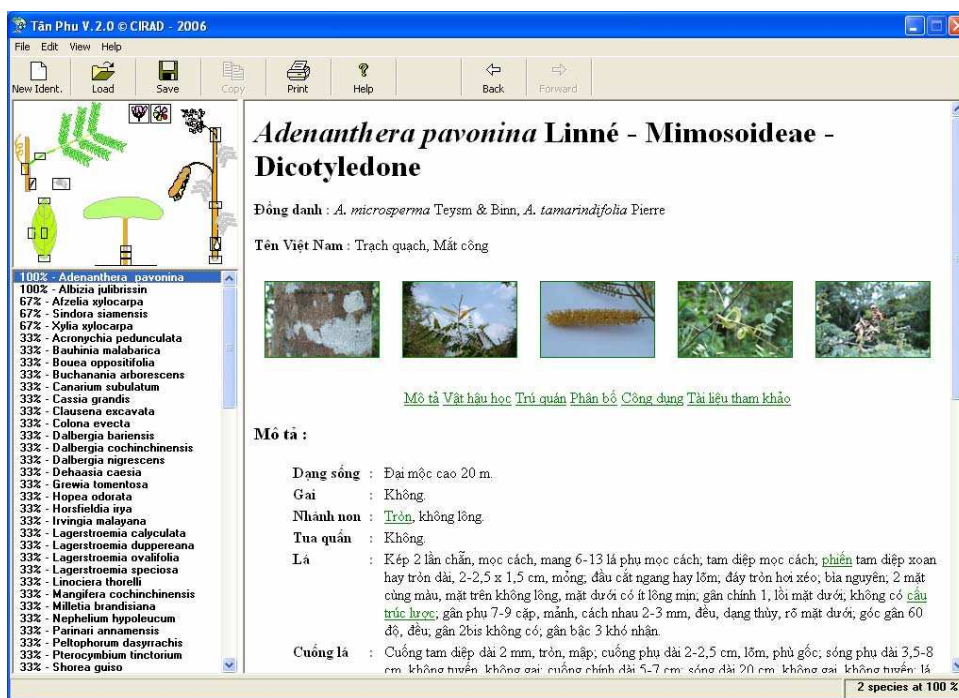
Hình 2. Giao diện của phần mềm Tân Phú

Với giao diện trực quan mô phỏng các đặc điểm nhận diện, người sử dụng (sinh viên cao đẳng, đại học ngành sinh học và các ngành có liên quan, cán bộ lâm nghiệp, cán bộ làm công tác bảo tồn, nhà nghiên cứu thực vật) nhanh chóng có được kết quả nhận diện ngay sau mỗi bước lựa chọn đặc điểm. Nguyên tắc của phần mềm là dựa trên đặc điểm đã được chọn để loại trừ dần các loài không phù hợp và hiển thị các loài tương thích. Chỉ với thao tác đơn giản là quan sát mẫu vật, người sử dụng so sánh và lần lượt chọn các hình mô phỏng trên giao diện của phần mềm sao cho phù hợp với đặc điểm của mẫu vật mà người sử dụng sẵn có hay quan tâm. Tùy vào trật tự, số lượng các đặc điểm được người sử dụng lựa chọn, phần mềm sẽ cung cấp kết quả tương ứng, có thể là 1 đến vài chục loài. Các loài này được sắp xếp theo thứ tự phần trăm tương thích giảm dần. Nếu kết quả tương thích có hơn 10 loài, người sử dụng có thể quay lại cửa sổ giao diện tiếp tục lựa

chọn đặc điểm để giảm số loài trong kết quả. Phần mềm cũng có biểu tượng “Tìm kiếm/Search” (Hình 2), giúp người sử dụng tiếp tục chọn lựa các đặc điểm khác để tiếp tục quá trình nhận diện. Còn trong trường hợp kết quả dưới 10 loài, người sử dụng có thể so sánh mẫu vật với các hình ảnh và mô tả của từng loài trong 10 loài này để tìm ra loài tương thích. Nếu có nhiều loài có cùng độ tương đồng thì phần mềm sẽ đề nghị các đặc điểm nên thu thập thêm (cũng thông qua biểu tượng "Tìm kiếm") để tiếp tục nhận diện các loài này. Trang kết quả của mỗi loài gồm bản mô tả chi tiết các đặc điểm, công dụng cũng như các hình chụp kèm theo. Trong phần mô tả đặc điểm thực vật của mỗi loài, người sử dụng có thể chọn đường dẫn để tìm hiểu định nghĩa có kèm hình ảnh minh họa các thuật ngữ chuyên môn. Một tiện ích khác của phần mềm là việc giúp người sử dụng nhận biết các lỗi gặp phải (biểu tượng “Errors”) trong quá trình nhận diện. Khi

trong trang kết quả của một loài hiện thị biểu tượng “Errors” có nghĩa đã có đặc điểm được chọn không phù hợp (được biểu thị bằng dấu X

đỏ trên cửa sổ giao diện. Khi đó, người sử dụng có thể quay lại cửa sổ giao diện để chọn lại đặc điểm khác phù hợp hơn.



Hình 3. Kết quả nhận diện tiếng Việt của phần mềm Tân Phú

KẾT LUẬN

Phát triển các phần mềm nhận diện thực vật là một xu thế đang phát triển tại Việt Nam. Dựa trên các mô tả hình thái, cơ sở dữ liệu và cấu trúc phần mềm IDAO, chúng tôi đang phát triển phần mềm nhận diện 265 loài cây gỗ rừng

Đông Nam Bộ, Việt Nam. Phần mềm này kế thừa các kết quả của các phiên bản có trước (Tân Phú, Cát Tiên), đồng thời được phát triển hơn về số lượng loài và các đặc điểm nhận diện, đặc biệt ở lá.

**IDENTIFICATION SOFTWARE OF 220 FOREST TREES IN SOUTHEASTERN
REGION, VIETNAM**

Dang Le Anh Tuan⁽¹⁾, **Jérôme Millet**⁽²⁾, **Nguyen Le Xuan Bach**⁽³⁾, **Tran Ninh,**
Pierre Bonnet⁽⁴⁾, **Pierre Grard**⁽⁴⁾

⁽¹⁾ University of Science, VNU-HCM

⁽²⁾ University Lyon1, France; ⁽³⁾ Institute of Tropical Biology

⁽⁴⁾ Botanic and Bio-informatic of plant architecture, France

ABSTRACT: *Using visual morphologic characteristics with more attention to leaf architecture and the structure of software IDAO created by CIRAD, we developed an identification software for 220 forest tree species of southeastern region, Vietnam, expected for 265 species in final version in 2011. By easily comparing and choosing the graphically diagnostic characteristics, users promptly achieve the results after each chosen feature. The results include botanical description with definitions of characteristics and the related pictures. If there are more suitable choices of species with the same similarity in the result, the software can suggest the additive necessary characteristics to continue the identification. Error display and modification of identifying process are two additional advantages of the software.*

Key words: IDAO, trees, Southeastern region Vietnam.

TÀI LIỆU THAM KHẢO

- [1]. Goodall, D.W. *Identification by computer*. Bioscience 18, 485-488 (1968).
- [2]. Hall, A.V. *A computer-based system for forming identification keys*. Taxon 19, 12-18 (1970).
- [3]. Lobanov A.L. *Results of experiments with biological diagnostic systems based on "Nairi-C" computer*. Syktyvkar: 162-168 (in Russian) (1975).
- [4]. Duncan, T, and Meacham, C.A. *Multiple-entry keys for the identification of angiosperm families using a microcomputer*. Taxon 35, 492-494 (1986).
- [5]. Dallwitz, M.J, Paine, T.A, and Zurcher, E.J. *Interactive identification using the Internet*. <http://delta-intkey.com> (2002).
- [6]. Harris J.G and Harris, M.W., *Plant Identification Terminology: Illustrated Glossary*. (2001).
- [7]. Phạm Hoàng Hộ. *Cây cỏ Việt Nam*, quyển I, II, III, NXB Trẻ (1999).
- [8]. Lecomte, H. *Flore générale de l'Indochine*, Tomes I-V, Paris (1938-1950).

- [9]. Aubréville, A. *et al. Flore du Cambodge, du Laos, et du Vietnam*, Fascicules 1-29, Paris (1960-1997).
- [10]. Phùng Mỹ Trung và cs. *Sinh vật rừng Việt Nam*. <http://www.vncreatures.net/> (xem ngày 30/04/2011)
- [11]. Đặng Lê Anh Tuấn. 2006. *Hình thái học ứng dụng: Phần mềm nhận cây gỗ rừng Tân Phú, tỉnh Đồng Nai*. Luận văn thạc sĩ. Trường Đại học Khoa học tự nhiên, Đại học quốc gia Tp.HCM. 110 trang