

Software de apoyo al proceso de identificación de especímenes

Estudio de las herramientas existentes

María Mora (INBio)



Contenido

- Objetivo
- Públicos meta prioritarios
- Requerimientos
- Tipos de herramientas existentes
- Metodología
- Demostración de algunas de las herramientas evaluadas
- Resumen de resultados
- Conclusiones

Objetivo

Conocer y evaluar las herramientas existentes para el apoyo de procesos de identificación de especímenes llevados a cabo por **usuarios no expertos** y seleccionar la que mejor se adapte a las necesidades de los usuarios (clientes) de INBio

Públicos meta prioritarios de INBio

- **Políticos:** formuladores de políticas.
- **Gestores de biodiversidad:** autoridades y personal técnico del MINAE, gestores de áreas protegidas públicas y privadas, cámaras, municipalidades.
- **Usuarios de recursos:** pescadores, turismo, otros empresarios.
- **Medios de comunicación y formadores de opinión**
- **Educadores y líderes religiosos**
- **ONG conservacionistas.**

Requerimientos

- Herramienta orientada a usuarios no expertos (altamente gráfica).
- Implementación de la funcionalidad en ambiente web.
- Que pueda ser integrada con las aplicaciones existentes.
- Que siga los lineamientos de la estrategia informática de la institución.
- Herramientas en proceso de desarrollo o en uso, con posibilidades de realizar o adiciones o extensiones.

Tipos de herramientas existentes

- Claves interactivas
- Inteligencia artificial
 - Sistemas basados en conocimiento.
 - Sistemas expertos
- Procesamiento de patrones
 - Sistemas de identificación automática por medio de imágenes.

Metodología

Definición de variables a evaluar

- 1. Implementación en ambiente web**
(Indispensable)
- 2. Manejo de información multimedial**
(Indispensable el uso de al menos imágenes)
- 3. Integrable** (Indispensable)
- 4. Fácil de usar** (Indispensable)
- 5. Escalable** (Indispensable)
- 6. Extensible** (Importante)
- 7. Costo** (Importante)
- 8. Documentación** (Importante)

Metodología

Definición de variables a evaluar (continuación)

9. **Compatible con estándares internacionales**
(De importancia media)
10. **Multilinguaje** (Indispensable)
11. **Eficiente** (Indispensable)
12. **Comunidad de usuarios** (Importante)
13. **Sostenibilidad** (de importancia media)
14. **Otras herramientas que apoyan al creador de la guía, clave o sistema experto y benefician al usuario final** (de importancia media)

Metodología

Definición de variables a evaluar (continuación)

- 15. Identificación con información incompleta**
(Importante)
- 16. Mecanismos para seleccionar caracteres prometedores.** (de importancia media)
- 17. Dependencias entre atributos.** (Importante)
- 18. Manejo de subclaves** (Importante)
- 19. Manejo de glosarios** (Importante)
- 20. Herramientas de software y protocolos utilizados en su desarrollo e implementación**

Metodología

Selección de las herramientas a evaluar

1. Recomendaciones internas.
2. Búsquedas en Internet
3. Sitio web de Delta

<http://delta-intkey.com/www/idprogs.htm>

4. Kew interactive

<http://kikforum.wordpress.com/tag/tools/>

5. Digital Taxonomy

http://digitaltaxonomy.infobio.net/index.php?Software:DELTA_and_Identification

6. Tesis y otras referencias

Inteligencia artificial

Area de investigación de las Ciencias de la Computación, cuyo objeto de estudio la *inteligencia*. El término *artificial* se refiere al intento por crear mecanismos inteligentes en un medio no biológico.

Inteligencia artificial

- **Sistema experto:** sistemas de IA, utilizados para resolver problemas de diferentes tareas de dominio (clasificación, planificación, diseño, diagnóstico, etc.).
- Generalmente rígidos, (basados en reglas), con capacidad limitada de aprendizaje.

Inteligencia artificial

- **Sistemas basados en conocimiento:** cuenta con una base de conocimiento inicial, provista por un experto.
- Recibe entradas expresadas en un lenguaje particular, y las manipula en un proceso de inferencia para adaptar la representación de su conocimiento, para proponer y explicar las soluciones encontradas.
- Aprende: transforma la representación de su conocimiento para mejorar su función de desempeño.

Sistemas de identificación automática por medio de imágenes.

- Smithsonian (IIS): Guía de campo electrónica
 - Biblioteca digital con información de especímenes (texto y fotos). 85K Tipos.
 - Algoritmo para reconocer plantas por medio de imágenes (hojas).
 - Identificación automática (dispositivos móviles).

IIS

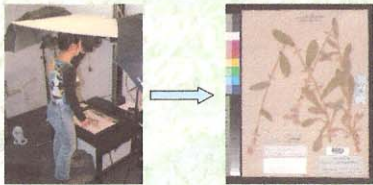
The Instant Identification System: Plant Exploration in the 21st Century

Department of Botany, National Museum of Natural History, Smithsonian Institution
Department of Computer Science, University of Maryland
Department of Computer Science, Columbia University

<http://herbarium.cs.columbia.edu>

In the Future

Type Specimen Digital Collection



- 84,000 Type Specimens and voucher specimens from
- the US Herbarium digitized and available online or on DVD
- vouchers from Plummers Island, MD and Washington-Baltimore area woody plants have been collected and digitized

Collection of Isolated Leaves



The database consists of 11,000 isolated leaves collected from the DC region.

Funded in part by NSF Grant IIS-03-25867. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views, opinions, or policy of the National Science Foundation.

Natural habitats are rapidly disappearing and many plant species may go extinct. Our purpose is to speed the pace of discovering and describing species by use of a computerized system for rapid and accurate field identification of living plants.

Goals of the project are to accelerate specimen identification by:

- photographing and electronically cataloguing herbaria
- making data available in the field
- assisting in locating the right information
- automating the search and match (ID) process
- building and testing a prototype field guide

IIS: The Process

1. Sample



Leaf image captured by wireless camera and transferred to tablet PC, where algorithm analysis then begins. Collected data is saved.

2. Capture



Sample automatically compared with leaves in the database based on shape via algorithm analysis.

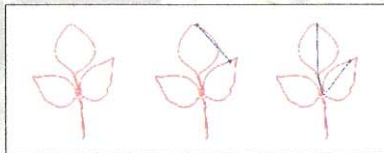
3. Determine



Specimen displayed, along with ranked results from matching algorithm. Text and images of matched species from the SI collection can be inspected.

Representations that Capture Part Structures

Euclidean IDSC*



* Inner Distance Shape Context (IDSC) – algorithm uses 64 points on a given leaf. Recognition rates are about 96% accurate.

Gathering Data

- using venation to ID leaves with nearly identical shape
- incorporating additional layers of information (text, photos, keys) to aid in the final ID
- develop web-based search and identification capability
- augmented reality prototypes
- field testing of tropical species

In Progress

The Future Field Biologist



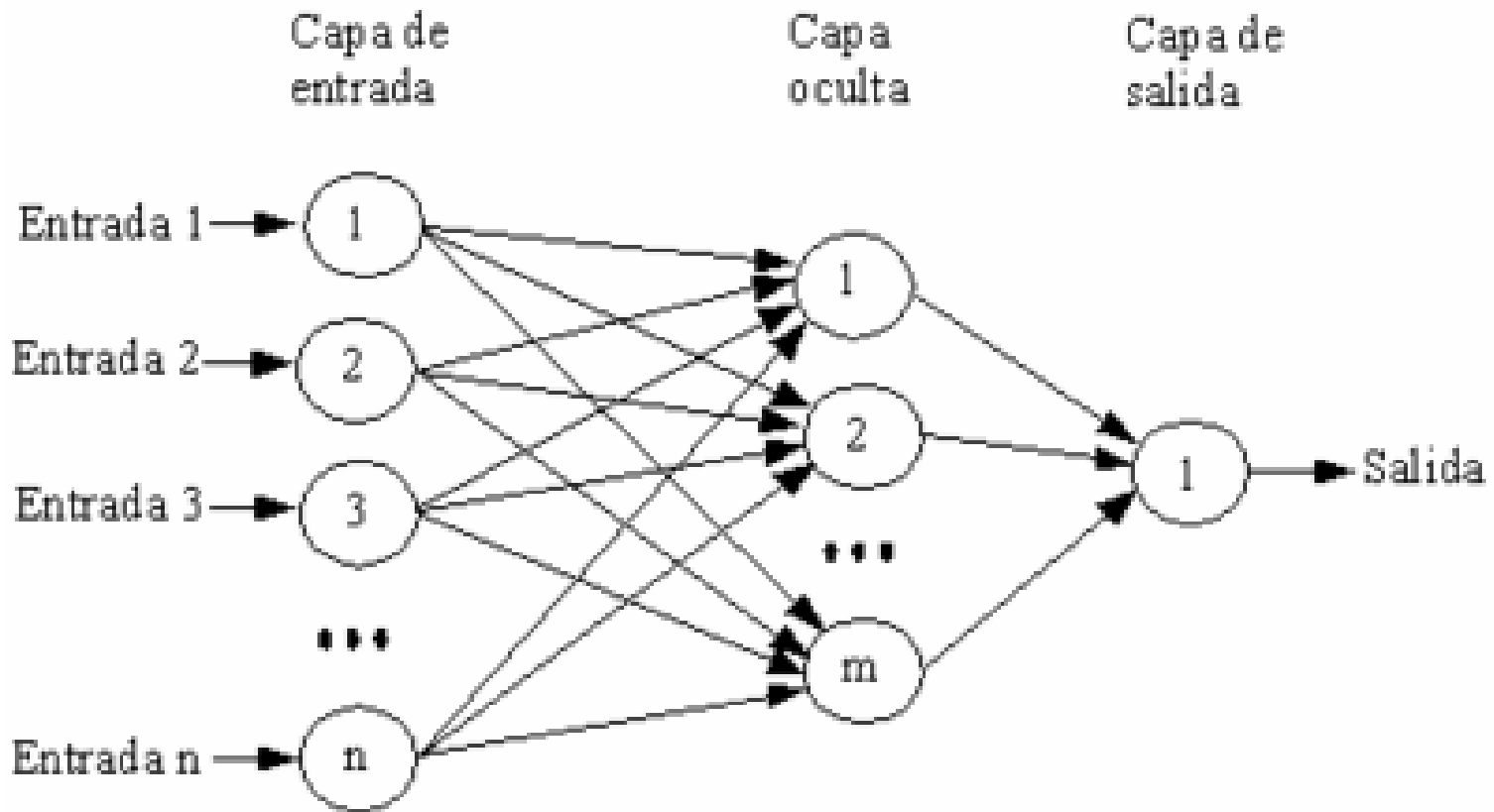
Future Uses:

- Biology
- Forensics
- Customs
- Border Entry Processing

Sistemas de identificación automática por medio de imágenes.

- DAISY (Digital Automated Identification System) <http://www.tdlegal.co.uk/>
 - Museo de Historia Natural de Londres.
 - Identificación de insectos.
 - No restringido a ningún tipo de organismo.
 - Red neural (Plastic self organising map (PSOM)) + correlación con el vecino más cercano.

Red Neuronal



Claves interactivas: Demostración de algunas herramientas

- [Lucid](#) (Universidad de Queensland)
- [Electronic Field Guides](#) (UMass y Missouri Botanical Garden)
- [INTKEY \(DELTA\)](#) (Agencia nacional de la ciencia, Australia)
- [X:ID](#) (Biblioteca del laboratorio marino de Woods Hole, Massachusetts)
- [ID Nature Guides \(Discover Life\)](#)
- [Linnaeus II \(IdentifyIt\)](#) (ETI)

Problemas de las claves

- **Obligan al usuario a buscar en una larga lista de atributos.**
- **Una sesión requiere de varios ciclos de proceso.**
- **En los sistemas guiados por árboles de decisión, el sistema pregunta "una cosa a la vez".**
- **Representación: los caracteres apropiados para un grupo pueden tener ninguna relación con otros grupos.**
- **Herramientas hechas por taxónomos orientadas a usuarios finales por lo general no son lo que los usuarios finales requieren.**

A thin green line starts horizontally from the left edge of the slide, extends across most of the width, and then slopes downward to the right edge.

Resumen de resultados

Conclusiones

- Los sistemas basados en conocimientos y sistemas expertos evaluados son experimentales o no son escalables.
- Los sistemas de reconocimiento de patrones deberían ser utilizados en combinación con otras herramientas.
- Las claves en estado avanzado de desarrollo y las más utilizadas son: DELTA – INTKEY, LUCID y IDNature Guides.

Conclusiones

- DELTA es un estándar y una herramienta. Es pionera en este campo. CSIRO detuvo el proceso de desarrollo. Ventaja: muchas herramientas leen el formato DELTA.
- LUICD y IDNature Guides son herramientas que pueden servir al proyecto.
- Electronic Field Guides (EFG) es una iniciativa nueva que está desarrollando herramientas para usuario final