



Date : 10/07/2006

**VIAF (Fichero de Autoridades Vitual Internacional)
Uniendo los ficheros de autoridad de persona de la
Deutsche Bibliothek y la Library of Congress**

Rick Bennett

OCLC Online Computer Library Center
Dublin, Ohio
USA

Christina Hengel-Dittrich

Die Deutsche Bibliothek
Frankfurt am Main
Germany

Edward T. O'Neill

OCLC Online Computer Library Center
Dublin, Ohio
USA

Barbara B. Tillett

Library of Congress
Washington, D.C.
USA

Meeting:	123 Cataloguing
Simultaneous Interpretation:	Yes

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL

20-24 August 2006, Seoul, Korea

<http://www.ifla.org/IV/ifla72/index.htm>

Abstract

La Deutsche Bibliothek, la Library of Congress, y OCLC están desarrollando conjuntamente un fichero de autoridades virtual internacional (VIAF) de nombres de persona que relaciona registros de autoridad de las agencias bibliográficas nacionales del mundo, y que estará libremente accesible en la red. Los objetivos del proyecto son comprobar la viabilidad de unir automáticamente registros de autoridad de diferentes ficheros nacionales y demostrar sus beneficios. Se utilizaron los ficheros de autoridad

y bibliográficos de la Library of Congress y la Deutsche Bibliothek para crear el VIAF inicial, que contiene más de seis millones de nombres con más de medio millón de links. Un aspecto clave del proyecto fue el desarrollo de algoritmos automatizados de comparación de nombres, que usan información de los registros de autoridad y de sus correspondientes registros bibliográficos. Se demostró lo práctico de unir los nombres de persona entre ficheros de autoridad nacionales mediante algoritmos. El 70 % de los registros de nombres de persona comunes a ambos ficheros fueron automáticamente relacionados con un porcentaje de error de menos del uno por ciento. El objetivo a largo plazo del proyecto VIAF es combinar las entradas de autoridad de muchas bibliotecas nacionales y otras fuentes significativas en un servicio de autoridades global compartido.

Traducción al español:
Ricardo Santos Muñoz
Biblioteca Nacional (España)

Introducción

Varios grupos dentro de la Sección de Catalogación de la IFLA reconocieron el potencial de crear un fichero de autoridades virtual internacional [1], donde los registros de autoridad de diferentes agencias bibliográficas nacionales del mundo que representen a una misma entidad fueran relacionados y estuviesen disponibles en Internet. Este VIAF sería una extensión práctica del concepto de control bibliográfico universal, y se construiría sobre el trabajo realizado por cada agencia bibliográfica nacional. Permitiría coexistir como formas autorizadas a las diferentes variantes nacionales o regionales, respetando las necesidades de los usuarios a nivel mundial en cuanto a preferencias de lengua o alfabeto.

Las propuestas de futuro actuales para Internet recogen el uso de ontologías para hacer una red más inteligente para el procesamiento automático. El VIAF podría ser uno de los elementos constitutivos para una “Web semántica” [2], combinada con otros vocabularios controlados y ficheros de autoridad provenientes de otras fuentes como servicios de indización y resumen, archivos, museos, editores, etc. Las bibliotecas tienen una oportunidad para aportar una gran contribución a este futuro, y deberían ayudar a hacer de esta visión una realidad. Es importante para esta visión compartida que el VIAF sea libremente accesible a los usuarios de todo el mundo.

Otros proyectos han considerado el enlazar nombres de persona en ficheros de autoridad. El Proyecto LEAF [3] (Uniendo y Explorando Ficheros de Autoridad) propuso enlazar registros de autoridad de diferentes fuentes, incluyendo bibliotecas, archivos y centros de documentación e investigación. Estos registros tienen formatos distintos, y el tipo y cantidad de contenido varían considerablemente. El proyecto LEAF propuso la relación automática de registros según son cargados en el sistema. Debido a las distintas fuentes de los registros de autoridad, descubrieron que la única información común para establecer estas uniones era el nombre, incluyendo las referencias de véase, y fechas asociadas. Dado que los registros de los participantes actuales frecuentemente no incluyen fechas, se espera que el porcentaje de error en la comparación sea inaceptablemente alto.

InterParty [4] es un proyecto de demostración financiado por la UE para crear ficheros de autoridad enlazados entre diversas organizaciones, cuyo principal propósito es apoyar la gestión de derechos digitales. Este sistema proporcionaría un único punto de acceso a las múltiples bases de datos involucradas en el sistema, suministrando un servicio centralizado de búsqueda. Como los links entre los nombres de las diferentes bases de datos son identificados manualmente, el mismo individuo que hace la asociación puede hacer el link, que ya pueden ser entonces usados automáticamente. Según la organización que haga los enlaces, así pueden ser considerados suficientemente fiables. La asunción de un link por alguna de las partes no significa que sea aceptado por las otras. El proyecto permite la posibilidad de comparación algorítmica, pero no se especifican las técnicas o requerimientos necesarios para permitir esta posibilidad.

El Proyecto VIAF

Durante el World Library and Information Congress de la IFLA de 2003 en Berlín, la Deutsche Bibliothek (DDB), la Library of Congress (LC) y OCLC Online Computer Library Center (OCLC), acordaron desarrollar un fichero de autoridades virtual internacional (VIAF) para nombres de persona [5] Los objetivos del proyecto VIAF son probar la viabilidad de enlazar automáticamente registros de autoridad de diferentes ficheros nacionales y demostrar los beneficios de un fichero VIAF. El proyecto VIAF unirá los ficheros de autoridad de persona de la LC y la DDB en un solo fichero de autoridad de nombres virtual. OCLC está desarrollando el software para comparar registros de autoridad de persona entre los dos ficheros de autoridad. El objetivo a largo plazo de VIAF es unir los nombres de autoridad de muchas bibliotecas nacionales y otras fuentes en un servicio de autoridades global para personas, entidades, conferencias, nombres geográficos, etc.

El proyecto VIAF consta de cinco fases:

1. Construcción de unos “registros de autoridad ampliados” procedentes de Personennormdatei (PND) y LC Authority Records. Estos registros ampliados contendrán una identificación de los registros correspondientes de autoridad e incluirán consideraciones sobre cualquier necesidad especial de manipulación de los registros entrantes.
2. Creación de la versión inicial del VIAF mediante el desarrollo de algoritmos de comparación y la comparación de los registros de autoridad ampliados de PND y LC. Fue un proceso iterativo con la fase 1, ya que pudo extraerse información adicional de los resultados intermedios de comparación e incluirse en el registro de autoridad ampliado para mejorar la comparación.
3. Construir un servidor OAI [6] para facilitar el acceso al VIAF
4. Mantenimiento de la base de datos VIAF, para lo que se requiere de las agencias participantes las adiciones y los cambios tanto de los registros de autoridad como de los bibliográficos. Esta tarea de actualización y mantenimiento será diseñada sobre los protocolos usados por OAI para las actualizaciones.
5. Para el acceso a los registros VIAF, se creará en la Web una interfaz de usuario. En un futuro, la base de datos y el interfaz soportarán Unicode y funcionalidades multi-lenguaje y multi-alfabeto. Para apoyar la Web semántica, se podrán usar peticiones directas a la base de datos, y se obtendrá, por ejemplo, un nombre de persona en su versión LC, y, con un simple link de HTML, su versión PND.

El proyecto inicialmente se centra en demostrar la idoneidad del VIAF uniendo los registros de autoridad de persona entre el Personennormdatei (PND) y el Library of Congress Name Authority File (LCNAF). A 31 de diciembre de 2005, el fichero LCNAF contenía 4.2 millones de registros de autoridad de persona. En la misma fecha, LC había creado y distribuido un total de 9.3 millones de registros bibliográficos.

En otoño de 2005, el fichero PND contenía 2,6 millones de registros de autoridad de nombres de persona. El fichero de autoridad PND se usa en los registros bibliográficos de la DDB y en los del Bibliotheksverbund Bayern (BVB). Entre ambos suman un total de 15 millones de registros bibliográficos asociados con los registros de autoridad PND.

El Problema de la comparación de nombres

En principio, el VIAF funcionará como un diccionario Alemán-Inglés, Inglés-Alemán para nombres de persona. Por ejemplo, si un usuario americano busca **J. P. De Valk** (la forma del nombre establecida por la LC), el nombre podría ser traducido automáticamente a **Johannes P. De Valk** (la forma establecida por DDB). Como se aprecia en este ejemplo, es algo común que diferentes agencias internacionales de catalogación establezcan diferentes nombres o, a la inversa, usen la misma forma representado autores distintos. Es posible que **J. P. de Valk** pudiera estar establecido por DDB para un autor completamente distinto.

Los nombres personales pueden adoptar distintas formas para la misma persona o tener la misma forma para diferentes personas, haciendo difícil la comparación de nombres procedentes de distintos ficheros de autoridad. La cobertura de ambos ficheros es significativamente diferente; solo una pequeña fracción de los nombres de persona están presentes en ambos ficheros. Por ello, además del nombre por sí mismo, se utilizan otros datos para asegurar la fiabilidad de la comparación. En registros de autoridad para nombres de persona, se halla a menudo presente la fecha de nacimiento o muerte del autor. La combinación de ambas fechas generalmente es suficiente para distinguir personas con nombres similares.

Para confirmar la dificultad de comparación de registros de autoridad que supone no usar esta información suplementaria, se extrajo de LC y DDB una muestra de nombres comunes. Estos pares de registros de autoridad fueron posteriormente revisados a mano para determinar si representaban a la misma persona. Como resultado se obtuvo que cerca del 10% de estas parejas de nombres representaban a personas distintas. Por consiguiente, el porcentaje de error usando solo la forma establecida del nombre sería inaceptablemente alto. Y dado que las formas del nombre no son siempre idénticas entre dos ficheros nacionales de autoridad, emparejar nombres similares, pero no idénticos, llevaría a un porcentaje de error aún más alto. Este sencillo enfoque también falla para comparar numerosos nombres que habían sido establecidos de forma diferente.

La solución al problema

Era evidente que se necesitaba información adicional para confirmar o rechazar posibles coincidencias de nombres. Por ejemplo, consideremos la siguiente información de autoridad de LC para Diane Glynn:

```
100 10 $a Glynn, Diane, $d 1946-
400 10 $a O'Connor, Diane, $d 1946- $w nna
670    $a Country western dancing, 1994: $b CIP t.p. (Diane
      Glynn) pub. info. (an avid country w. dancer & co-author
      of How to make your man more sensitive)
```

El único dato utilizable directamente son el nombre y la fecha de nacimiento. En el campo 670 (Fuente consultada con datos) se citan dos obras que pueden ser extraídas y procesadas. En la práctica, sólo algunos de las obras pueden ser extraídas con fiabilidad.

Una fuente obvia de información adicional sobre la persona son los registros bibliográficos. Estos registros pueden ser analizados para extraer atributos adicionales sobre el trabajo de una persona, que puede servir para distinguir a una persona de otras con nombres similares. Uno de los registros bibliográficos es así:

```
100 1 $a Glynn, Diane, $d 1946- -  
245 10 $a How to make your man more sensitive / $c by Diane and  
      Dick O'Connor.  
700 1 $a O'Connor, Dick, $d 1938- $e joint author -
```

Los registros bibliográficos contienen 2 tipos de información adicional. Normalmente contienen información específica de la obra como el título, e información específica de la manifestación como el número de ISBN. Una comparación de títulos supone un dato casi definitivo para una identificación de nombres.

El registro bibliográfico contiene igualmente información adicional de aplicación para obras en colaboración del autor. Esta información puede ser de utilidad para comparar nombres cuando no sea posible la comparación de título anteriormente mencionada. El autor en colaboración Dick O'Connor es un ejemplo de esto. Dick O'Connor puede ser coautor de más de un libro con Diane Glynn, lo que es un elemento importante para identificar nombres en diferentes ficheros de autoridad. Incluso si la misma obra aparece en ambas bases de datos nacionales, la identificación de los nombres puede ser difícil de realizar automáticamente, si en una de ellas se encuentra traducida. En este caso, la autoría conjunta es mucho más posible que ocurre en ambas bases de datos, confirmando la identidad.

Todos los registros bibliográficos disponibles donde el nombre es un encabezamiento principal, secundario o materia, son transformados para crear un registro intermedio llamado "autoridad derivada". Estos registros de autoridad derivada se combinan entonces con el registro original de autoridad para crear el registro de autoridad ampliado. Debido a que estos registros ampliados incluyen información adicional asociada con el nombre procedente de los registros bibliográficos, el proceso de comparación e identificación es mucho más riguroso que los registros de autoridad por sí mismos.

Confirmación de la identificación del nombre

Comparar simplemente los nombres de dos ficheros nacionales de autoridad puede ser un modo razonable de localizar al mismo individuo. Ya que se pueden dar formas variantes del mismo nombre, se reduce la posibilidad de que estos individuos sean la misma persona. Para confirmar automáticamente la identificación de estos individuos, el procedimiento seguido aquí es, 1) los nombres deben ser compatibles, y 2) debe existir información complementaria suficiente que confirme la identificación.

La compatibilidad requiere que no existan diferencias que invaliden que los nombres representen a la misma persona. Los nombres pueden variar entre formas más o menos completas, como John A. Smith y John Allen Smith. Estos nombres son compatibles porque "A" podría significar "Allen". Pero John A. Smith y John B. Smith no son compatibles porque existe un conflicto entre las 2 iniciales intermedias. Ambas formas, la autorizada y las variantes, son tenidas en cuenta a la hora de comprobar la compatibilidad.

Una vez que se ha determinado la compatibilidad de los nombres, se emplea la información adicional recogida de estos nombres para confirmar su identidad. Los ficheros bibliográficos pueden contener muchos títulos y muchos nombres diferentes pero similares. Si un par nombre/título es similar en ambos ficheros es muy posible que ambos nombres representen a la misma persona. Esta estrategia básica se extiende también a otros tipos de información extraída de los registros bibliográficos.

Las fechas se consideran separadamente, en una correlación positiva. Cuando dos fechas difieren por más de un año, los nombres se consideran incompatibles, y la identificación se rechaza. Se permitieron discrepancias en una fecha de un año: durante el desarrollo del VIAF, fue relativamente común encontrar pequeñas discrepancias en algunas de las fechas, y la restante información adicional fue suficiente para confirmar la identificación, incluso con estas ligeras diferencias en las fechas.

Al comparar dos registros de autoridad ampliados, cada elemento que coincide es considerado como un punto de coincidencia. Estos puntos se separan en tres categorías: fuerte, moderado y débil. Para nombres compatibles, un punto fuerte de coincidencia es considerado como suficiente para confirmar que los individuos son la misma persona. Los puntos de coincidencia fuertes son títulos, ISBN, fechas de nacimiento y muerte, o relación de coautoría. La mera fecha de nacimiento no fue suficiente para diferenciar nombres, y se consideró como un punto moderado. Los puntos de coincidencia moderados son indicadores del entorno de trabajo de las personas, tales como el editor, área temática o la función del autor (ilustrador, compositor...). Una gran compañía editorial publica las obras de muchos autores, y al menos algunos de ellos pueden tener nombres parecidos. Las coincidencias de múltiples puntos moderados son suficientes para confirmar la identificación. Los puntos débiles son considerados suficientes únicamente para diferenciar identificaciones que de otro modo serían ambiguas. Ejemplos de estos puntos de coincidencia débiles serían lengua, área temática y país de publicación.

Para combinar los puntos de coincidencia, a cada uno de ellos se le asigna una puntuación numérica. Para un número, como el ISBN, la coincidencia es exacta o no lo es, y por ello se le asigna una puntuación de 1 o 0, según si es coincidente o no. Para un dato textual, como un título, se le asigna un número, según su semejanza, entre cero y uno. La puntuación de la semejanza de un dato textual se asigna mediante una técnica de puntuación basada en trigramas. La puntuación individual se modifica por un peso basado en su importancia (fuerte, moderado o débil), y se suma. Cuando la puntuación total supera un límite determinado durante el proceso de comprobación, la identificación se confirma. En el algoritmo de comparación, la comprobación de muchos registros permitió el ajuste de la puntuación dentro de estas categorías, y se espera que los ajustes continúen a medida que se añadan más registros al sistema y se gane experiencia.

Construyendo los “registros de autoridad ampliados “

Las técnicas descritas anteriormente fueron las utilizadas para crear los registros de autoridad ampliados en los ficheros de PND y LC. Los ficheros bibliográficos de LC fueron procesados para crear registros de autoridad derivados y ampliar los registros de autoridad de LC, y los bibliográficos de DDB y BVB lo fueron para ampliar los

registros de autoridad de PND. La Figura 1 muestra un breve diagrama del flujo de información que permitió los registros ampliados.

Para el fichero de LC, 3.8 de 4.2 millones (90%) de registros de autoridad pudieron ser ampliados. Solo 2.6 millones (60%) fueron ampliados con información extraída de registros bibliográficos, un total de 7.4 millones de títulos. Otras ampliaciones se hicieron con los 4.1 millones de títulos de obras extraídos del campo 670 (Fuente consultada con datos) de los registros de autoridad. Los títulos son el elemento de ampliación más importante, como se verá en la sección de resultados.

Para el fichero de PND, 2.4 de 2.6 millones (90%) registros de autoridad recibieron alguna ampliación, aunque sólo 2 millones (80%) lo fueron gracias a registros bibliográficos. Los restantes 400000 registros fueron ampliados con títulos provenientes de los propios registros de autoridad de PND.

Técnicas de comprobación de coincidencias

Los participantes de VIAF colaboraron al proceso de comparación mediante revisiones de la precisión y el comentario de los resultados. Por ejemplo, los títulos de serie se usaron al comienzo, pero se comprobó que a menudo se identifican nombres incorrectamente. Cada revisión daba como resultado modificaciones que, o bien aumentaban el número de identificaciones, o disminuían el número de identificaciones erróneas. Durante ese tiempo, se desarrolló un algoritmo y un nivel de puntuación razonablemente preciso. Aquí sólo se describen los resultados de la comprobación final.

Con el objetivo de comprobar la precisión y efectividad del proceso de comparación, un equipo de bibliotecarios con experiencia de DDB y LC analizó un muestreo del proceso. El primer muestreo tenía dos objetivos, determinar el grado de elementos comunes en ambas bases de datos, y descubrir qué fracción de esas parejas de nombres comunes se identifican en el proceso de comparación. El segundo muestreo se utilizó para detectar errores del sistema o deficiencias que pudieran ser subsanados y estimar el porcentaje total de error.

La primera muestra contenía 391 registros de PND seleccionados aleatoriamente. Se buscaron en la base de datos de LC, manual y automáticamente, coincidencias de estos registros. Mediante búsqueda automática, se emparejaron los registros de autoridad de PND con todos los registros de LC que compartieran un apellido, resultando 74000 parejas para su examen. Se aplicó a este conjunto el algoritmo de comparación y se identificaron automáticamente 79 pares de registros de autoridad PND/LC.

La revisión manual de los 391 registros de PND aportó 35 nombres más que tenían una correspondencia en LC, pero, o bien la correspondencia no se basaba en un apellido, o el algoritmo de comparación falló a la hora de confirmar la identificación. Se confirmó que las 79 correspondencias realizadas automáticamente eran correctas. Usando la muestra de PND, se estimó que alrededor del 30% de los nombres de PND aparecen también en la base de datos de LC, y que el algoritmo es capaz de identificar alrededor de un 70% de estos nombres comunes. Esto se traduce en unos 800000 nombres en común entre los dos ficheros de autoridad de los cuales se espera que el proceso automático sea capaz de identificar 550000.

Los resultados también fueron revisados con el fin de mejorar el proceso de emparejamiento. Usando solamente apellidos, se determinó que por cada identificación realizada, necesitarías un proceso completo de comparación unos 100 pares de nombres. El resultado del test manual se usó para determinar que una estrategia basada en el apellido, nombre de pila, y una somera información de fechas, podría dar una estimación aproximada de la compatibilidad de nombres. Este índice simple localizó el 95% de estas coincidencias, con sólo cuatro pares examinados por cada identificación realizada. El índice simple es eficiente y efectivo, y probablemente con unos ligeros retoques se conseguirían aún mejores resultados.

El objetivo del segundo muestreo era estimar la tasa de error. Como parte del proceso, el muestreo puso a prueba la idoneidad del límite de puntuación marcado para establecer la identificación entre los pares. Cuando se trabaja con estos límites de puntuación, el porcentaje de error de las puntuaciones que quedan cerca de este umbral es mayor que aquellas que lo sobrepasan ampliamente. La mayoría de las identificaciones de autoridades tenían puntuaciones muy por encima del valor establecido como límite. Para obtener la mejor estimación de la tasa de error con el menor trabajo manual posible la muestra fue dividida en cuatro sub-muestras, basadas en la puntuación. Las revisiones manuales detectaron los errores, y para cada sub-muestra se estableció una tasa de error y fiabilidad. Estos resultados parciales fueron ponderados y sumados para obtener el porcentaje general de error para el procedimiento de comparación. El número de identificaciones erróneas resultó ser menor del uno por ciento.

Uno de los subconjuntos quedó justo bajo el límite de aceptación. Si se bajaba este límite, una identificación incorrecta pasaría por cada tres correctas. Se consideró claramente injustificado bajar el umbral. En el rango de puntuación justo encima del límite, había solamente una correspondencia errónea por cada 25 correspondencias establecidas. Dado que en este rango hay relativamente pocas correspondencias, el impacto total en el porcentaje de error es bajo, y se logra un alto número de correspondencias. Por todo ello, el primer límite que se estableció fue el que se dio como válido.

Construyendo el VIAF inicial

Los registros de autoridad ampliados de ambas fuentes pasaron a través del algoritmo de comparación, y los registros resultantes, los coincidentes y los no coincidentes, se convirtieron a registros VIAF. El proceso se ilustra en la Figura 2. Existen 6.3 millones de registros en el archivo VIAF incluyendo 500000 registros enlazados, 3.7 millones de registros sin ningún enlace del LC, y 2.1 sin enlace de PND. El dato está muy cercano al que se estimó en la comprobación manual. Se estima que existen además unos 250000 pares de registros de autoridad que representan a la misma persona que no pudieron ser automáticamente enlazados debido a falta de información útil. El sistema final permitirá la asociación manual entre estos registros, así como el establecimiento de otras asociaciones intelectuales entre ellos. Los registros de autoridad incluirán un número de registro VIAF asignado secuencialmente.

En la Figura 3 se observa un ejemplo de registro VIAF en formato MARC21. Dado que el primer objetivo de VIAF es establecer una unión entre los registros, el registro VIAF contiene una entrada por cada nombre en el campo 700, junto con una indicación de su

fuelle. Al no haber un único nombre autorizado el campo 100 (Encabezamiento de nombre de persona) no se utiliza. Cuando, mediante el algoritmo, se determina una coincidencia, en el registro se añaden dos enlaces. Cuando para un nombre no existe ninguna coincidencia, solo aparecerá un único campo 700.

La información suplementaria también se incluye en los registros de autoridad ampliados, como campos locales (9xx). En la Figura 4 se describen brevemente los campos locales usados en los registros de autoridad ampliados. Para simplificar la comparación, el texto está normalizado mediante una versión modificada de las reglas de normalización NACO (Programa Cooperativo de Autoridades Personales del Programa de Catalogación Cooperativa). [7] El número de ocurrencias de un término particular se almacena en el subcampo \$9. Este número no se visualizará necesariamente en la versión para el usuario, ya que es un dato pensado para el procesamiento automático. A medida que se sumen nuevos ficheros nacionales de autoridad, se compararán estos datos con los ya existentes en los registros VIAF ampliados, y se incorporarán los resultados según se vayan añadiendo al VIAF los registros. Cuando se realicen los nuevos enlaces, la información ampliada de los nuevos registros será igualmente añadida.

En un número de casos significativo, un nombre autorizado de uno de los ficheros equivale a múltiples nombres autorizados en el otro fichero. Como el objetivo de VIAF es un esquema de relaciones uno a uno, en estos casos de equivalencia múltiple, no se realiza el enlace, por lo que 70000 coincidencias algorítmicas fueron eliminadas. Se detectaron al menos dos razones que explican esta equivalencia múltiple.

En primer lugar, hay una serie de nombres no diferenciados en el PND, cada uno equivalente a dos o más nombres diferenciados en el LCNAF. Siguiendo las reglas de catalogación alemanas RAK-WB, la práctica catalográfica en Alemania era no diferenciar nombres personales. Cuando la DDB empezó a catalogar usando el fichero de autoridad, se abandonó esta práctica, y hoy la DDB ya no crea registros de autoridad de persona no diferenciados. Sin embargo, el PND contiene aún muchos nombres no diferenciados. DDB diferenciará estos nombres con equivalencia múltiple, automáticamente hasta donde sea posible, usando como base los títulos consignados en los registros ampliados de LC y DDB; el resto se hará intelectualmente. Estas correcciones se harán presentes en el VIAF como parte de sus frecuentes actualizaciones y proporcionarán enlaces sin ambigüedades entre los registros comparados.

En segundo lugar, los registros de autoridad de LC reflejan la práctica de AACR2 de mantener registros de autoridad separados para cada identidad bibliográfica usada por una persona, como ocurre con los seudónimos. Este es el caso contrario a los registros no diferenciados de PND. En este caso, existen múltiples registros de autoridad para un solo individuo. El PND, siguiendo las reglas RAK-WB, tiene solo un registro de autoridad para todas las identidades. Como con los nombres no diferenciados, estos registros de autoridad “sobre-diferenciados” presentan problemas para los cuales no ha sido aún encontrada una solución satisfactoria.

Los registros enlazados pueden ser utilizados directamente como un traductor automático de las autoridades de LC a las de PND y viceversa. Se contribuye así a los requerimientos de la Web semántica o sistemas de búsquedas federadas que requieren

esta característica. El mantener los "véase además" puede también proporcionar información adicional al usuario.

Los números identificadores de los registros autoridad participantes o los propios números de registros VIAF, pueden ser usados como base para URIs, pudiendo potencialmente convertirse en un servicio de gestión de URIs de autoridades. Partiendo de una cita de un URI que aparezca en un documento, registro o página Web, un usuario podría navegar hacia todos los materiales, registros, recursos, etc., con los que una autoridad, representada por el URI, está relacionada, y hacia los propios registros de autoridad.

Mantenimiento del sistema

Los ficheros nacionales de autoridad y las bases de datos bibliográficas cambian constantemente. Con una base de datos enlazada construida sobre dos o más ficheros que cambian constantemente, dichos enlaces deben ser reevaluados y modificados constantemente. La lógica y el software que soporta el sistema VIAF inicial se está modificando para permitir la continua actualización de los registros. Según se reciben nuevos registros bibliográficos o de autoridad, se modifican los registros de autoridad ampliados existentes, y se reevalúa la base de datos cruzada. Constantemente se realizarán nuevas comparaciones y enlaces, y se destruirán las identificaciones que no deban mantenerse en función de cambios en las fuentes que las soportan. Cuando se destruyan, en cada registro se mantendrá, para referencia, un histórico de relaciones previas para referencia

En el futuro, el sistema VIAF se alimentará de las bases de datos participantes mediante OAI, a medida que estén disponibles. Entretanto, se usarán métodos de acceso más tradicionales, como FTP.

Con una gran cantidad de datos en una única localización, se pueden diseñar una gran cantidad de modalidades de acceso. Los enlaces se pueden usar para traducir un nombre personal a un formato final elegido por el usuario, como se requiere en la Web semántica. Se pueden desarrollar herramientas para realizar búsquedas automáticas en diferentes bases de datos, para que en cada una de ellas se busque en la forma del nombre apropiada para cada base de datos. Se pueden desarrollar del mismo modo herramientas de catalogación, identificando la forma apropiada de un nombre incluido en el registro bibliográfico. Y por supuesto, se podrá buscar directamente en la base de datos VIAF.

Conclusiones

El archivo PND ya se benefició sustancialmente del proyecto. Los tests de comparación en ambos ficheros supusieron importantes mejoras en PND, y la DDB espera obtener a través de la comparación de títulos en los pares de registros ampliados, una ayuda sustancial para la diferenciación de nombres de persona. Los procesos y algoritmos de comparación desarrollados para el proyecto son adaptables para otras aplicaciones. Se están investigando utilidades que aprovecharían los datos de comparación de nombres de persona para mejorar el acceso a información bibliográfica y servir de ayuda a las actividades de catalogación de los participantes.

El proyecto ha demostrado que enlazar automáticamente los nombres de persona de dos ficheros de autoridad nacionales es práctico. El 70 % de los registros comunes fueron enlazados con una tasa de error de menos del uno por ciento. La estrategia de complementar los registros originales de autoridad con información de los registros bibliográficos incrementó enormemente el porcentaje de identificación al mismo tiempo que disminuyó el de las identificaciones erróneas. Unos pequeños cambios en los registros de autoridad hubieran mejorado significativamente el proceso. Muchos fallos vinieron de errores al tratar el campo 670 (Fuente con datos). Sería muy útil una estructura adicional, evitando los títulos y nombres abreviados, o enlaces explícitos a la fuente bibliográfica. También el identificar explícitamente el papel habitual o especialidad de los autores (compositor, ilustrador, matemático, etc.) así como incluir las formas completas de los nombres, aunque fuera como referencias cruzadas, ayudaría al proceso de comparación.

La investigación respalda el uso del control de autoridades y el uso de registros de autoridad, la utilidad de las referencias cruzadas y en red, y la construcción de Web semánticas para bibliotecas. Para aquellas bibliotecas y redes de bibliotecas de Alemania que obtienen o crean registros bibliográficos con puntos de acceso de LCNAF, el VIAF podría servir como plataforma para cruzar de un fichero de autoridad al otro, bien para transcribir los puntos de acceso de LCNAF en registros bibliográficos como puntos de acceso PND, o para permitir buscar y recuperar con encabezamientos PND en el VIAF. En portales multinacionales o multilingües como el de la European Library, el VIAF podría automáticamente combinar búsquedas en LCNAF y PND al mismo tiempo, para que el usuario obtenga registros bibliográficos de ambas fuentes.

Con las técnicas de comparación empleadas, se piensa en un sistema actualizable que recogería datos de autoridad y bibliográficos actuales de nombres de persona usando el protocolo OAI. El sistema está diseñado para ser escalable, por lo que serían bienvenidos nuevos participantes dispuestos a compartir sus registros de autoridad y bibliográficos. Los límites de la escalabilidad del VIAF no podrán determinarse hasta que no se incorporen nuevas instituciones al proyecto.

El proyecto VIAF se ha centrado en el problema de comparar registros de autoridad. Para mantener, expandir e implementar el VIAF, se necesitará una estrategia de gobierno y servicio a largo plazo. Se necesita igualmente tomar decisiones sobre la expansión del proyecto para incluir nombres corporativos y para la inclusión de instituciones participantes. Existen planes para expandir la capacidad del sistema incorporando el juego de caracteres Unicode. Unicode permitirá la inclusión de nuevos alfabetos, pero la expansión del algoritmo de comparación supondrá un reto, particularmente para alfabetos ideográficos como el coreano, el chino o el japonés.

Referencias

1. IFLA Core Activity: IFLA-CDNL Alliance for Bibliographic Standards (ICABS) <http://www.ifla.org.sg/VI/7/icabs.htm> [May 2006]
2. Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities." *Scientific American*, May 17, 2001. <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> [May 2006]
3. LEAF Project, <http://www.leaf-eu.org> [May 2006]
4. Project InterParty: From Library Authority Files to E-Commerce, Andrew MacEwan, http://www.haworthpress.com/store/E-Text/View_EText.asp?a=3&fn=J104v39n01_11&i=1%2F2&s=J104&v=39 [May 2006]
5. VIAF: The Virtual International Authority File, <http://www.oclc.org/research/projects/viaf> [May 2006]
6. Open Archives Initiative - Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html> [May 2006]
7. Hickey, Thomas B., Jenny Toves, and Edward T. O'Neill. "NACO Normalization: A detailed Examination of the Authority File Comparison Rules", *Library Resources & Technical Services*, Vol. 50, No. 3, p. 18-24. [forthcoming]

Figura 1. Creación del registro de autoridad ampliado

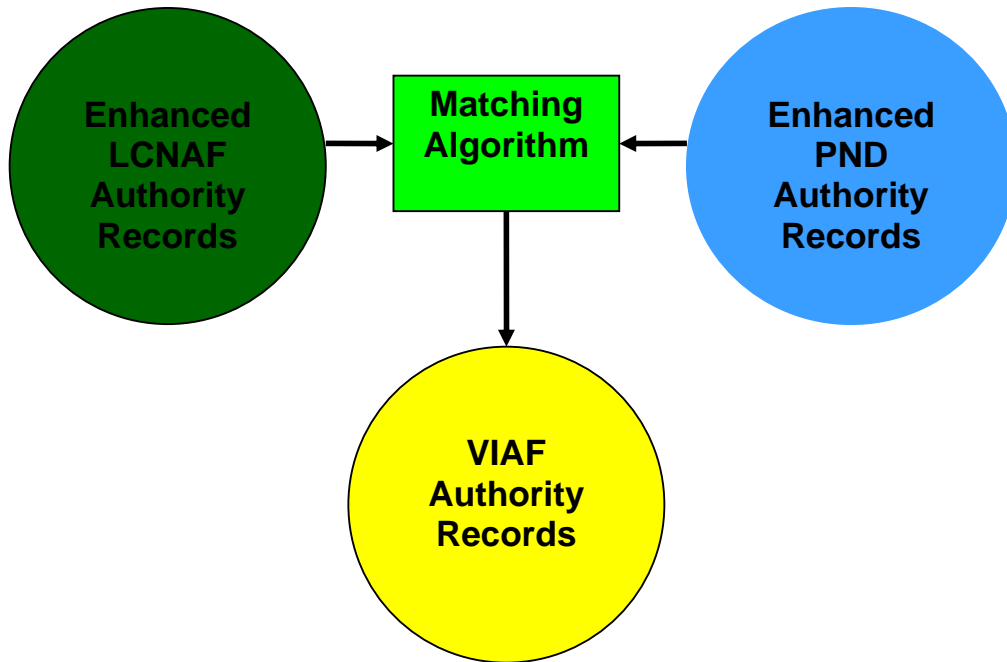


Figura 2. Creando el registro de autoridad VIAF

```
000    nz n
001    viaf 30543
005    20050826163535.0
008    050826n||anannabbn |a aaa
040    VIAF    $c VIAF
400 10    $w nnaO'Connor, Diane,    $d 1946-
700 17    Glynn, Diane,    $d 1946-    $2 DLC    $0 n 94057411
700 17    O'Connor, Diane    $2 DDB    $0 108982424
901    052512920    $9 1
901    349917275    $9 1
901    350215532    $9 1
903    75014386    $9 1
910 11    how to make your man more sensitive    $9 3
910 11    macht eure manner zartlicher    $b liebevolle ratschlage
fur e neues rollenverhalten    $9 1
910 11    macht eure manner zartlicher    $b wie e frau ihrem mann
helfen kann e verstandnisvoll    $9 1
919    country western dancing,    $9 1
920    0-525    $9 1
920    3-499    $9 1
```

920	3-502	\$9	1	
921	dutton	\$9	1	
921	rowohlt	\$9	1	
921	schertz	\$9	1	
922	gw	\$9	2	
922	nyu	\$9	1	
940	eng	\$9	1	
940	ger	\$9	2	
942	18	\$9	1	
943	197x	\$9	3	
944	am	\$9	3	
950	11	oconnor, dick	\$9	2
950	11	oconnor, dick	\$d 1938	\$9 1
999	1	\$b 75014386 //r94	\$2	DLC
999	1	\$b n 94057411	\$2	LoCNA
999	2	\$b 780147766 \$b 790425319	\$2	DDB

Figura 3. Registro VIAF

Figura 4
Formatos del registro ampliado

90x Números de control

901	ISBN	\$a Parte numérica del ISBN (sin dígito de control ni guiones)
902	ISSN	\$a Parte numérica del ISSN (sin dígito de control ni guiones)
903	LCCN	\$a Parte numérica del LCCN (sin dígito de control ni guiones)

91x Campos de título

910	Título del 245 Título abreviado	Subcampos 'a' y 'b'
911	del 210 Título uniforme del	Subcampos 'a' y 'b'
913	130 or 240 Título traducido	Subcampos 'a' y 'b'
914	del 242 Título uniforme	Subcampos 'a' y 'b'
915	colectivo del 243 Variante de título	Todos los subcampos
916	del 246 Título uniforme del	Subcampos 'a' y 'b'
917	registro de autoridad Título extraído de	Extraído de registros de autoridad de autor/título, campo 100 \$t
919	otra fuente	Notas o campos similares

92x Campos de publicación

920	Número del editor	\$a Número del editor del ISBN
921	Nombre del editor Lugar de	\$a Nombre del editor del 260 b or 533 c.
922	publicación	\$a Código del país de publicación del 008

93x Uso

930	Forma del nombre usada	\$a Forma del nombre como aparece en la mención de responsabilidad del 245 subcampo 'c'
-----	------------------------	---

94x Atributos

940	Lengua	\$a Código de lengua del 008 o 041 \$a
941	Función del autor	\$a Código del campo 700 \$e y/o 4
942	Materia NATC Década de	\$a Número NATC
943	publicación	\$a Década de publicación
944	Formato	\$a Tipo y nivel bibliográfico (008 06-07)
945	Materia Conspectus	Según PND
95x Co-autores		
		Subcampos \$a, \$b, \$c, \$d, y \$q de los campos 100 o 700
950	Autores personales	
951	Autores corporativos	Subcampo \$a de los campos 110 o 710
96x Nombre como materia		
	Nombre como	
960	materia	Sub-campos \$a, \$b, \$c, \$d, y \$q del campo 600 Texto "Materia" que indica que el encabezamiento fue usado como materia, extraído del campo 600
969	Uso como materia	
99x Campos especiales		
	Registros	\$a Número total de registros
	bibliográficos	\$b Número de control del registro
999	asociados	\$2 Fuente del registro