

Reconocimiento de Imágenes basado en el Contenido de Baladas Impresas en Volantes: La Herramienta ImageMatch de las Bibliotecas Bodleianas

Traducción al español del documento original en inglés: "Content-Based Image-Recognition on Printed Broadside Ballads: The Bodleian Libraries' ImageMatch Tool"

Traducido por: Dirección de Traducciones de la Biblioteca del Congreso de la Nación Argentina (traducciones@bcn.gob.ar)

Buenos Aires, República Argentina

Giles Bergel, Facultad Lengua y Literatura Inglesas, Universidad de Oxford, Oxford, Reino Unido. giles.bergel@ell.ox.ac.uk

Alexandra Franklin, Bibliotecas Bodleianas, Universidad de Oxford, Oxford, Reino Unido. alexandra.franklin@bodleian.ox.ac.uk

Michael Heaney, Bibliotecas Bodleianas, Universidad de Oxford, Oxford, Reino Unido.
michael.heaney@bodleian.ox.ac.uk

Relja Arandjelovic, Departamento de Ingeniería, Universidad de Oxford, Oxford, Reino Unido.

Andrew Zisserman, Departamento de Ingeniería, Universidad de Oxford, Oxford, Reino Unido.

Donata Funke, Bayerische Staatsbibliothek, Munich, República Federal de Alemania.



Copyright © 2013 by **Giles Bergel, Alexandra Franklin, Michael Heaney, Relja Arandjelovic, Andrew Zisserman** and **Donata Funke**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

URI: <http://library.ifla.org/id/eprint/209>

Resumen

Este trabajo introduce la herramienta ImageMatch de las Baladas archivadas en la Biblioteca Bodleiana, desarrollada por el Grupo de Geometría Visual del Departamento de Ingeniería de Software de la Universidad de Oxford por encargo de las Bibliotecas Bodleianas. ImageMatch fue diseñada como herramienta de apoyo para la catalogación y el estudio del contenido pictórico de baladas británicas antiguas impresas en volantes, pero podría utilizarse también para muchos otros tipos de material impreso. El trabajo describe la naturaleza de los materiales a los cuales se ha aplicado ImageMatch, el modo en que la herramienta funciona y lo que puede hacer, y analiza los beneficios de ImageMatch para catalogar imágenes en colecciones de Libros Poco Comunes.

Palabras Clave: Bodleiana, baladas, volantes, grabados sobre madera, reconocimiento de imágenes

Este trabajo introduce la herramienta ImageMatch de las Baladas Bodleianas, desarrollada por el Grupo de Geometría Visual del Departamento de Ingeniería de Software de la Universidad de Oxford por encargo de las Bibliotecas Bodleianas. ImageMatch fue diseñada como herramienta de apoyo para la catalogación y el estudio del contenido pictórico de baladas británicas antiguas impresas en volantes, pero podría utilizarse también para muchos otros tipos de material impreso. El trabajo describe la naturaleza de los materiales a los cuales se ha aplicado ImageMatch, el modo en que la herramienta funciona y lo que puede hacer, y analiza los beneficios de ImageMatch para catalogar imágenes en colecciones especiales y de libros poco comunes.ⁱ

Las baladas en volantes son hojas impresas de bajo costo con letras de baladas, ilustraciones y nombres de melodías populares. Se vendieron, exhibieron, cantaron y leyeron en las calles y tabernas británicas entre los siglos 16 y principios del 20.ⁱⁱ



Fig.1: Ejemplo de una balada del siglo 17 impresa en un volante (Bodleian MS Wood E25 (1))

Además de su contenido literario y musical, las baladas en volantes son especímenes de cultura visual. Se imprimían en un solo lado de la hoja para poder exhibirse públicamente y, a menudo, contenían ilustraciones hechas con impresiones de grabados sobre madera o, más tarde, con xilografía. Las ilustraciones podían acompañar la letra de una determinada canción, mostrando una imagen del tema de la balada, o un incidente dentro de la narrativa. En otros casos, las ilustraciones se relacionaban con otras ilustraciones de la misma hoja o eran trabajos independientes. En ocasiones, no existía una relación clara entre el texto y la ilustración, y esta se elegía para equilibrar la impaginación de la hoja, para transmitir una cualidad genérica o para expresar un sentimiento. iii

Se imprimían literalmente millones de copias de baladas en volantes, pero solo han sobrevivido unas pocas. Sin embargo, la Biblioteca Bodleiana cuenta con más de 34.000 ejemplares de baladas en volantes, en colecciones que datan del siglo 17 al 20. La mayoría de las baladas está disponible en línea desde hace quince años, pero en los últimos 18 meses la Biblioteca Bodleiana ha trabajado para reemplazar la base de datos utilizando un nuevo modelo de catalogación relacional que incorpora todos los metadatos previos, imágenes y otros contenidos, junto con muchas mejoras, incluida la herramienta ImageMatch.

Broadside Ballads Online (beta)

from the Bodleian Libraries



Home Browse Collections Help About Style:

Collection Statistics

35254	Ballads
30864	Sheets
29639	People
22717	Illustrations
420	Locations

Search

e.g. Hard times come again no more

Broadside Ballads Online presents a digital collection of printed songs made between the 16th and 20th centuries, linked to other resources for study of the English ballad tradition.

This resource is maintained by the Bodleian Libraries and features the Bodleian digital collections of ballads, with links to the English Broadside Ballad Archive's digital presentations of pre-1800 ballads from other libraries, and to the folk song scholarship of the Roud Broadside Index, hosted by the English Folk Dance and Song Society.

Featured Ballad

Harding B 11(650)

This sheet contains two ballads: the minstrel song *Nely Bly* sits incongruously with the stirring patriotic song *The Colours the Ladies Should Wear*.

<http://ballads.bodleian.ox.ac.uk>

Fig. 2: Baladas Bodleianas En Línea

Dada su naturaleza multimedia compleja, las baladas en volantes son difíciles de describir. Muchos rasgos típicos de las baladas presentan desafíos a la práctica angloamericana tradicional de catalogación: la importancia de la primera línea de la balada para establecer la identidad del texto; la ausencia de un título general para una hoja de baladas que puede contener más de una canción; la usual falta de detalles de impresión (lugar y fecha de publicación, nombre de la imprenta); la cantidad variable de baladas en una sola hoja; la importancia central de sus ilustraciones y la reutilización de componentes de una hoja en otras hojas. La Biblioteca Bodleiana ha utilizado una serie de soluciones de catalogación para adaptarse a la complejidad de los materiales y a las necesidades de los investigadores, que pueden ser especialistas en historia social, historia de la música folclórica, cultura popular, ilustración, música, impresión o venta de libros en el Reino Unido. Se ha publicado un nuevo modelo de catalogación relacional como esquema RDF que está disponible para ser utilizado y adaptado.^{iv} La primera base de datos en línea de baladas en volantes de las colecciones bodleianas, creada entre 1995 y 1999, implementó un índice de materias de las ilustraciones, basado en el sistema de indexación iconográfica ICONCLASS.^v



Fig. 3: Ejemplo del buscador ICONCLASS de Baladas Bodleianas

ICONCLASS es más detallado que otras herramientas como el Tesoro de Materiales Gráficos de la Biblioteca del Congreso, pero los esquemas de clasificación requieren una clave o código para comprender y utilizar el indexador que asigna los mismos códigos a situaciones idénticas. Esto puede ser muy difícil cuando se clasifica una colección grande.

A diferencia de lo que sucede con la descripción y el análisis iconográficos de ICONCLASS, ImageMatch no ofrece una categorización de las imágenes por tema, sino que está diseñada para ayudar al catalogador y al usuario a localizar ilustraciones derivadas de la misma plancha xilográfica u otra superficie de impresión.



Fig. 5: Plancha xilográfica de una balada (Museo Británico 2000, 0723.9) y su correspondiente impresión

Hacer coincidir planchas xilográficas es una habilidad interesante para catalogadores, bibliógrafos e historiadores del arte por una serie de motivos. El uso de una plancha xilográfica identificable permite asignar una fecha a la impresión de una hoja de baladas. El estado en el que se encuentra una plancha xilográfica, según lo revelan rasgos de la impresión, como agujeros de gusanos u otros daños, puede ayudar al catalogador a asignar una fecha o rango de fechas a las impresiones de una hoja de baladas dentro de una secuencia de impresiones.^{vi} El uso de una plancha xilográfica identificable puede revelar la imprenta de una hoja anónima de baladas, su lugar de impresión, la venta o préstamo de una plancha o la existencia de una relación entre imprentas.

Identificar planchas comunes es una tarea exigente que puede verse complicada por distintos factores: las variaciones en la colocación de la tinta, las fallas en las planchas y las hojas, la resolución, la modificación según una escala y la inclinación de la imagen digital sustituta, y la presencia de muchas copias sobre temas populares, que en algunos casos pueden ser difíciles de distinguir. A continuación mostramos un ejemplo de muchas versiones de una misma imagen, impresas a partir de diferentes planchas:

Versions of a similar image



Fig. 6: variaciones de una misma ilustración

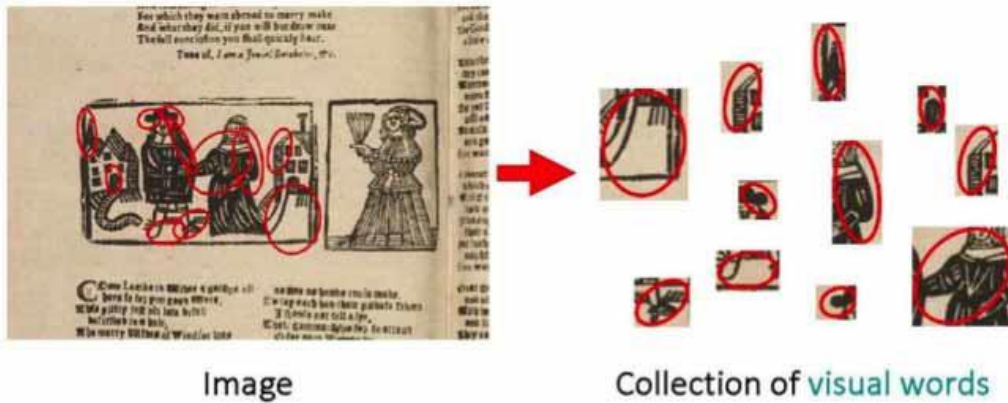
ImageMatch ofrece dos métodos para ayudar al catalogador a cotejar planchas. Estos métodos sirven también para distinguir copias obtenidas a partir de planchas diferentes. Los dos modos de funcionamiento son: el modo de recuperación de imágenes y el modo de comparación de imágenes.

ImageMatch está disponible para recuperar imágenes en el recurso “Bodleian Ballads Online” (Baladas Bodleianas En Línea), en forma integrada con los metadatos referidos a las colecciones. También está disponible como herramienta independiente. Se ha implementado en una subserie de unas 800 imágenes de baladas de alta resolución del siglo 17, cuando los volantes solían ilustrarse con una serie común de grabados sobre madera. Se activa seleccionando una región de la imagen de una hoja de baladas de la colección, subiendo un archivo de imagen, o ingresando el URL de una imagen archivada en otro lugar.

Las tareas de búsqueda de correspondencia y recuperación se realizan mediante el tratamiento de una imagen como un grupo de ‘palabras visuales’ (parches icónicos o fragmentos) que están unidas por elipsis de distintos tamaños. El software realiza una búsqueda entre las palabras visuales contenidas en la imagen seleccionada comparándolas con un índice pre-confeccionado de imágenes de la base de datos considerado como una colección de palabras visuales de idéntica constitución, utilizando un algoritmo de búsqueda eficiente similar a la búsqueda por palabra clave.viii

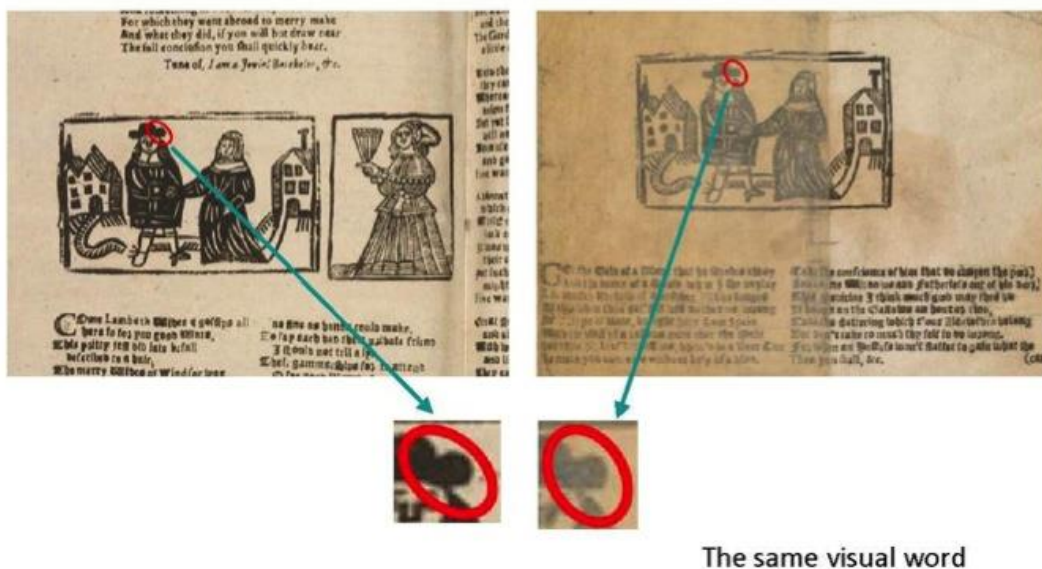
How it works

- Representation: bag of (visual) words
- Visual words are 'iconic' image patches or fragments
 - represent the frequency of word occurrence
 - but not their position

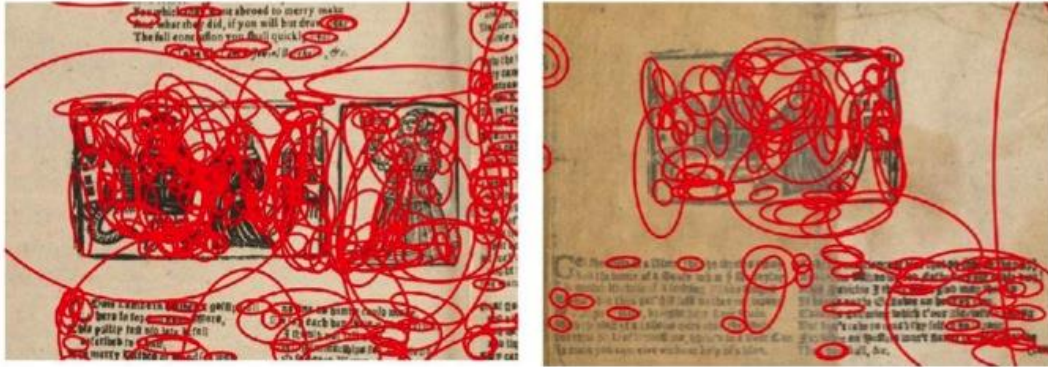


Visual vocabulary unaffected by printing artifacts

- Differences in printing strength, spatial deformations



Representation: bag of (visual) words



Use efficient Google like search on visual words

Fig. 7, 8, 9: Búsqueda por 'Palabras Visuales'.

El sistema muestra imágenes coincidentes clasificadas por grado de similitud, que se expresa mediante un puntaje numérico. El puntaje más alto corresponde a la primera imagen, es decir, a aquella que es idéntica.



Numerical scores indicate similarity of matching

Fig. 10: los puntajes numéricos indican el grado de similitud

Verificamos la precisión comparando los resultados de una muestra de imágenes (en formatos tanto bitonales como de alta resolución y con imágenes cargadas e imágenes guardadas en la base de datos como consultas) con la información comprobada en el terreno (*ground-truth*) obtenida mediante la identificación manual de planchas e ilustraciones similares.

La siguiente imagen muestra el funcionamiento del software al cotejar cuatro planchas, dos de las cuales son copias muy parecidas.

lavern men		Test 1	Judgement of Paris								village couple		Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	
		v											v	v	v	v	v	v	v	v	v	v	v
hit nr			hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr	hit nr		
4o Rawl. 566(132)	17		0/1	8	6	7	8	2	6	7		0/1	3	3	10	4	3	8	6	10	10		
4o Rawl. 566(135)	2		8	0/1	8	8	6	7	8	8		2	0/1	2	5	2	2	8	7	8	8		
4o Rawl. 566(137)	8		3	3	0/1	2	2	3	2	2		3	4	0/1	9	3	4	10	10	7	7		
4o Rawl. 566(141)	18		5	4	2	0/1	3	4	4	3		8	6	8	0/1	8	7	4	5	5	4		
4o Rawl. 566(151)	19		8	2	3	3	0/1	5	3	4		5	5	5	8	0/1	5	7	9	8	8		
4o Rawl. 566(187)	11		2	5	7	6	7	0/1	7	8		4	2	4	7	5	0/1	9	5	9	9		
4o Rawl. 566(199)	0/1		4	8	4	4	4	6	0/1	5		8	10	9	3	6	8	0/1	3	4	3		
4o Rawl. 566(34)	14		7	7	5	5	5	8	5	0/1		7	8	9	4	9	8	2	0/1	2	2		
4o Rawl. 566(46)	20		score: 14.000000		8							10	9	10	6	10	10	5	4	0/1	5		
4o Rawl. 566(84)	16		score: 13.000000		8							8	7	7	2	7	8	3	2	3	0/1		
Douce Ballads 1(122a)	21		score: 37.000000			8																	
Douce Ballads 1(55a)	4		score: 32.000000				8																
Douce Ballads 1(69a)1	7		score: 33.000000					8															
Douce Ballads 1(71a)	12		score: 24.000000						8														
Douce Ballads 1(86a)	9		score: 25.000000							8													
Douce Ballads 2(143a)	8		score: 20.000000								8												
Douce Ballads 2(160a)	10																						
Douce Ballads 2(180a)1	3																						
Douce Ballads 2(212a)	22																						
MS. Wood E 25(150)	5																						
MS. Wood E 25(58)	15																						
MS. Wood E 25(63)	13																						

ImageMatch testing

Fig. 11: Verificación de la precisión de ImageMatch con información comprobada en el terreno

En resumen, podemos observar que ImageMatch encuentra más de un 90% de imágenes relacionadas en una prueba con diecisiete planchas divididas en cuatro grupos con distintos grados de similitud.

Grupo prueba	de	Cantidad pruebas	de	Cantidad planchas	de	Precisión promedio	Recuperación promedio
1		14		2		0.66	1
2		3		2		0.69	0.86
3		12		5		0.96	1
4		7		8		0.72	0.85
TOTAL		36		17		0.76	0.93

Fig. 12: Tabla de precisión de ImageMatch y porcentajes de recuperación

Si bien ImageMatch está diseñada para cotejar impresiones tomadas de planchas enteras ordenadas según su fidelidad, la clasificación también puede incluir copias de planchas muy parecidas, incluso cuando se ha copiado solo una parte de la plancha. A continuación podemos observar dos planchas distintas pero que incluyen copias del recuadro externo y algunas similitudes en la figura del centro.

Similar images

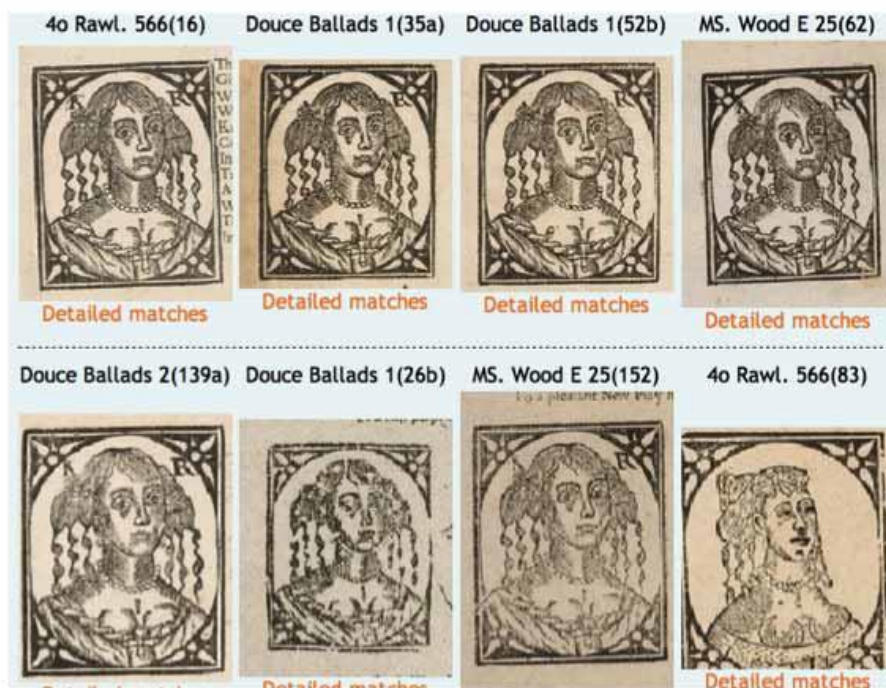


Fig. 13: Copias de planchas con elementos similares

También puede rastrear imágenes fragmentadas en las que se ha dañado una plancha o una impresión.

Fragmented images



Fig. 14: Recuperación de detalles de ilustraciones incompletas (dañadas)

Además, puede cotejar imágenes compuestas formadas por fragmentos de planchas que aparecen en otras combinaciones.

Fragmented and recomposed blocks



Fig.15: Fragmentos de planchas.

Esta versatilidad es beneficiosa para la catalogación iconográfica, al rastrear las tradiciones visuales y exhibir al mismo tiempo la complejidad de una cultura de la xilografía en la que las planchas xilográficas o sus elementos se intercambiaban, dividían, reformaban y copiaban.

El modo de comparación de imágenes se utiliza para un análisis más detallado de posibles coincidencias de planchas. Al igual que los compaginadores bibliográficos ópticos, que se basan generalmente en la superposición o estereoscopia, el modo de comparación de imagen proporciona una superposición, o registro, entre dos imágenes con el efecto de resaltar las diferencias.^{viii}

El registro se produce por una simple transformación bidimensional. Esto puede compensar la mayoría de las diferencias de escala presentadas por la fotografía. La identificación de planchas o láminas no es un procedimiento bien documentado. La mayoría de los investigadores se basa en la experiencia para determinar las coincidencias entre impresiones: Un área distintiva de daño (como un agujero de gusano) es la fuente más común de evidencia positiva de una correspondencia, mientras que una discrepancia significativa entre las dimensiones de una impresión, o entre series de puntos dentro de las impresiones, es la fuente más fehaciente de ausencia de coincidencia. Sorprendentemente, el ojo humano suele verse abrumado por una profusión de puntos de datos, complicados por las variaciones de tinta, el deterioro por uso y la posibilidad de una impresión calcada o una plancha copiada.^{ix} Todos estos obstáculos hacen que sea difícil llegar a una conclusión en un período corto de tiempo.

El modo de comparación de imagen puede, en primer lugar, reducir el tamaño de posibles planchas comunes descartando las discordancias evidentes. Luego puede ayudar al catalogador a determinar la existencia de imágenes más similares provenientes de la misma plancha.

La base de datos original de baladas Bodleianas incluía un índice iconográfico de las ilustraciones. El laborioso proceso de trabajo quince años atrás incluía el análisis de cada volante a nivel de edición y la descripción de las ilustraciones, haciendo lo posible por repetir las descripciones cuando se sabía que las ilustraciones eran similares. Llevó 2 años crear entradas de índice para 22.800 ilustraciones. La búsqueda de coincidencias entre imágenes acelera este proceso. Si al describir una imagen logramos encontrar e identificar inmediatamente imágenes idénticas y similares, podemos asignarles a todas los mismos códigos en forma segura. Por ejemplo, las 20 versiones de esta escena de hombres conversando alrededor de la mesa pueden haber sido codificadas en forma simultánea.

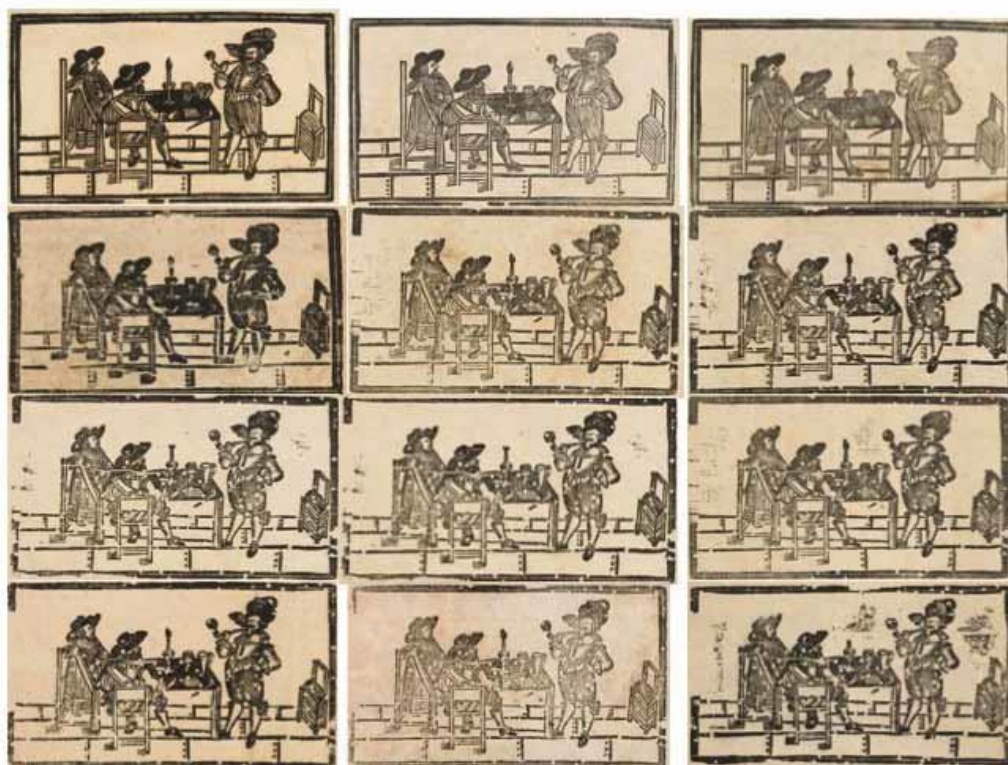
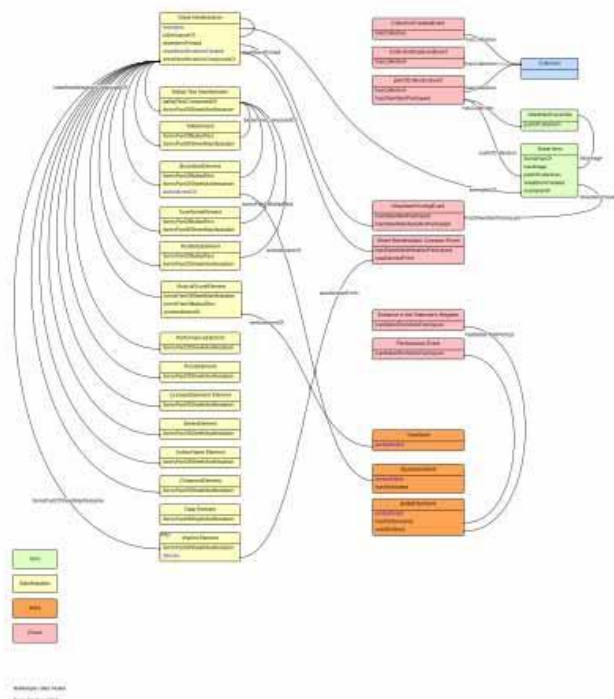


Fig.16: Múltiples impresiones de la misma plancha

Lo que no se intentó en 1999 fue realizar un catálogo que diera una identidad a cada plancha xilográfica.^x Ahora que es posible cotejar imágenes para agrupar impresiones, es posible crear una identidad para la matriz de imagen (en este caso, la plancha xilográfica utilizada). Esto simplificará mucho el proceso de creación de índices de ilustraciones para cualquier colección de obras impresas. Las imágenes coincidentes obtenidas de la misma plancha xilográfica ahora pueden ser agrupadas desde el comienzo y se les puede asignar una identidad. Una vez que se ha identificado una imagen, sus coordenadas en la hoja pueden agregarse también al registro del catálogo, y

pueden asociarse con los códigos ICONCLASS correspondientes y otros descriptores presentes en el registro del catálogo.^{xi} De esta manera, el primer nivel de descripción, que simplemente explica de qué se trata la imagen, puede realizarse de una vez para todas las impresiones de ese grabado sobre madera. La búsqueda de coincidencias entre imágenes que encuentra imágenes similares pero no idénticas también es útil e implica un ahorro de tiempo para el catalogador: las descripciones pueden ser replicadas para obtener varias versiones de una imagen, o modificadas cuando se descubren imágenes ligeramente diferentes. Finalmente, la explicación de lo que las ilustraciones podrían significar en relación con el texto de una balada constituye otro nivel de descripción, y una vez que las imágenes han sido adecuadamente identificadas, los catalogadores pueden encarar este paso con la seguridad de que la descripción inherente de las Imágenes sigue estando vinculada a sus identidades. Todos estos aspectos de la catalogación de ilustraciones, con respecto al ítem, de la manifestación, con respecto a la plancha, y los aspectos relativos al trabajo iconográfico, están comprendidos en el nuevo modelo de catalogación de baladas.



Diapositiva 17: Balladspec: un esquema RDF para baladas en volantes.

ImageMatch tiene un gran potencial de reutilización en una amplia gama de imágenes impresas. Puede aplicarse fácilmente a otras imágenes impresas en planchas xilográficas de origen similar: a nivel experimental, las imágenes dentro del corpus de baladas indexadas han sido cotejadas exitosamente con las imágenes obtenidas de otras colecciones digitales, como Early English Books Online (EEBO) (Libros Ingleses Antiguos En línea). Los elementos gráficos de impresiones tipográficas excepto las planchas xilográficas, como los ornamentos de imprenta, reglas, iniciales e incluso las

características menores como los diferentes tipos de letra también se pueden recuperar. Es probable que ImageMatch y otras herramientas similares de reconocimiento de imagen se conviertan rápidamente en una característica común de las bases de datos de imágenes impresas, ya sea en un sistema federado o en índices por actividad para clases específicas de materiales.^{xii} Sin embargo, las expectativas de los usuarios de un grado de precisión similar al que puede obtenerse a través de la búsqueda de texto pueden ser difíciles de satisfacer debido a la amplia gama de elementos gráficos presentes en el corpus de impresión, los diferentes estados de preservación y la calidad de las imágenes.^{xiii} Otros procesos de impresión, como el grabado sobre madera, el entallado o los métodos litográficos que producen un rango tonal más amplio que las impresiones tipográficas de alto contraste, plantean desafíos diferentes como, por ejemplo, qué sistemas de recuperación de imágenes deberían modificarse.

ImageMatch es tal vez la herramienta de catalogación más valiosa para un corpus cerrado de materiales. Debido a la creciente indexación de distintos corpus, su aporte será cada vez más valioso -a través de la función de carga- para el tratamiento de los materiales que aún no han sido descritos. Para estos materiales, puede ser útil tanto para la catalogación de su contenido gráfico como para la catalogación por copia de un ítem completo. En el futuro se analizarán los límites de la actual tecnología a través de la experimentación con la indexación a gran escala de los materiales ya digitalizados y a través de estudios más detallados de corpus específicos. Los conservadores, catalogadores, desarrolladores técnicos y usuarios finales tendrán un rol importante en el desarrollo de su potencial.

i ImageMatch funciona actualmente en los servidores de la Biblioteca Bodleiana como parte del recurso “Bodleian Ballads Online” en <http://ballads.bodleian.ox.ac.uk>, desarrollado como parte del proyecto “Archivo Integrado de Baladas en Volantes”, financiado por el Consejo de Sistemas de Información del Reino Unido (http://www.jisc.ac.uk/whatwedo/programmes/digitisation/content2011_2013/broadsideballads.aspx). También está alojado en el Visual Geometry Group (Grupo de Geometría Visual) junto con las especificaciones técnicas, los tutoriales y el código fuente (<http://www.robots.ox.ac.uk/~vgg>). La historia del proyecto, incluyendo videos demostrativos, puede encontrarse en <http://balladsblog.bodleian.ox.ac.uk/blog/570>.

ii La bibliografía sobre baladas en volantes es amplia: para un resumen de la forma, ver Leslie Shepard, *The Broadside Ballad (Baladas en Volantes)* (Herber Jenkins, 1962) y Claude Simpson, *The British Broadside Ballad and Its Music (La Balada Británica en volantes y Su Música)* (Rutgers University Press, 1966). Además de Bodleian Ballads Online (<http://ballads.bodleian.ox.ac.uk>), otras colecciones digitales incluyen el “English Broadside Ballad Archive” (Archivo de Baladas Inglesas en Volantes) (<http://ebba.english.ucsb.edu>) de la Universidad de California en Santa Barbara y “Word on the Streets” (La palabra en las calles) de la Biblioteca Nacional de Escocia (<http://digital.nls.uk/broadsides/index.html>). Los índices de Roud de la Sociedad Inglesa de Música y Danza Folclórica (<http://library.ehdss.org/cgi-bin/textpage.cgi?file=aboutRoud>) vinculan el contenido de las baladas en volantes con el de las baladas en otros formatos o medios.

iii Ver Alexandra Franklin, ‘The Art of Illustration in Bodleian Broadside Ballads before 1820’ (El Arte de la Ilustración en las Baladas Bodleianas en Volantes antes de 1820), *Bodleian Library Record* 17 (2002): 327-352

iv El esquema RDF será publicado en vocab.ox.ac.uk.

v El buscador ICONCLASS [<http://www.iconclass.nl/iconclass-2100-browser>] organiza los temas iconográficos en jerarquías y les asigna un código en un formato que aumenta la longitud del código en relación con la especificidad del tema. Por lo tanto, un código corto de dos dígitos indica un tema general, como “Naturaleza” o “Sociedad”, mientras que las letras y números agregados especifican: “palmera” o

“concierto musical”. Con el uso del buscador ICONCLASS, la búsqueda jerárquica desde categorías generales hasta ejemplos específicos puede ser simplificada por los hilos de metáfora y referencia que agrupan a los ítems en diferentes categorías. Esto permite la búsqueda cruzada de palabras clave vinculadas a los códigos, de modo que una "flecha" puede aparecer como un atributo del dios Cupido o como el arma de un soldado, en diferentes partes de la jerarquía.

vi Para agujeros de gusano, ver SB Hedges, ‘Wormholes record species history in space and time’ (Historia de especies de agujeros de gusano en el espacio y el tiempo), *Biological Letters*, 9 (2012)

vii Ver R. Arandjelovic and A. Zisserman, ‘Three things everyone should know to improve object retrieval’ (Tres cosas que todos debemos saber para mejorar la recuperación de objetos), *IEEE Conference on Computer Vision and Pattern Recognition*, 2012; J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, ‘Object Retrieval with Large Vocabularies and Fast Spatial Matching’ (Recuperación de Objetos con Vocabularios Amplios y Correspondencia Espacial Rápida), *IEEE Conference on Computer Vision and Pattern Recognition*, 2007

viii El Registro (superposición) como herramienta de apoyo para la detección de diferencias en impresiones puede lograrse con un programa estándar de procesamiento de imágenes, con el uso de transparencias fotográficas, o con dispositivos ópticos de recopilación (para estos últimos, ver Steven Escar Smith, ‘The Eternal Verities Verified:

Charlton Hinman and the Roots of Mechanical Collation’ (Las Verdades Eternas Comprobadas: Charlton Hincan y las Raíces de la Recopilación Mecánica), *Studies in Bibliography*, 53, (2000). 129-161 y la documentación para el recopilador McLeod mencionada en

<http://blogs.bodleian.ox.ac.uk/theconveyor/2010/09/03/library-machines-the-mcleod-collator>

ix Las planchas pueden ser copiadas, ya sea en forma directa o a partir de sus impresiones, con un alto grado de precisión. Para la copia de impresiones, ver W. M. Ivins, *How Prints Look: Photographs with a Commentary*, (Apariencia de las Impresiones: Fotografías con Comentarios)(New York Metropolitan Museum of Art, 1943), 117-41. Para la copia y reproducción de planchas en metal, ver James Mosley, ‘Dabbing, abklatschen, clichage...’ en http://typefoundry.blogspot.co.uk/2006_01_01_archive.html

x Sin embargo, los catalogadores pueden haber advertido la existencia de diferentes planchas, en particular a los fines de identificar diferentes estados, cuestiones o ediciones a los cuales pertenece el ítem en su conjunto. Un recurso que distingue a las planchas xilográficas dentro del registro de impresiones es *A Guide to English Illustrated Books*(Una Guía de Libros Ingleses Ilustrados) 1536-1603, Ruth Samson Luborsky and Elizabeth Morley Ingram (Medieval and Renaissance Texts and Studies, 1998). Para la copia, venta y reutilización de planchas xilográficas, ver Barry McKay ‘Cumbrian Chapbook Cuts: Some Sources and Other Versions’ (Panfletos de Cumbria: Algunas Fuentes y Otras Versiones), en Peter Isaac and Barry McKay (eds.), *The Reach of Print: Making, Selling and Using Books* (El Alcance de la Imprenta: Producción, Venta y Utilización de Libros) (St. Paul's Bibliographies, 1998) 65-83

xi Del total de 24.405 ediciones de las colecciones Bodleianas, 15.919, o el 65%, contienen ilustraciones. La cantidad total de ilustraciones, teniendo en cuenta que algunas ediciones contienen más de una imagen, es 22.800.

xii El reconocimiento de imágenes basado en el contenido para imágenes impresas también se ha implementado en el Banco Internacional de Ornamentos de Imprenta de Passe-partout (<http://www3.unil.ch/BCUTodai/>); en el buscador Japonés de impresiones en Planchas Xilográficas <http://ukiyo-e.org>; y en el buscador de imágenes basadas en la similitud de la Biblioteca Estatal de Baviera <http://bildsuche.digitalesammlungen.de>

xiii La mayoría de las imágenes de Baladas Bodleianas son escaneos bitonales provenientes de microfilms: esta serie de imágenes ha resultado ser bastante menos susceptible de ser recuperada con ImageMatch, en particular cuando se selecciona una imagen de gran calidad para analizarla con un índice de menor calidad. Se requiere más investigación y desarrollo en materia de adaptación de dominio para mejorar la búsqueda de coincidencias en colecciones de imágenes heterogéneas y definir estándares óptimos de imágenes para el futuro.

