

User Perceptions of Relevance and Its Effect on Retrieval in a Smart Textile Archive

Ben Horsburgh^{1,2}, Susan Craw^{1,2}, Dorothy Williams³,
Simon Burnett³, Katie Morrison³, and Suzanne Martin⁴

¹ School of Computing Science & Digital Media

² IDEAS Research Institute

³ IMAGeS Research Institute

Robert Gordon University, Aberdeen, Scotland, UK

{b.horsburgh,s.craw,d.williams,k.morrison}@rgu.ac.uk

⁴ School of Textiles & Design

Heriot-Watt University, Galashiels, Scotland, UK

s.e.martin@hw.ac.uk

Abstract. The digitisation of physical textiles archives is an important process for the Scottish textiles industry. This transformation creates an easy access point to a wide breadth of knowledge, which can be used to understand historical context and inspire future creativity. The creation of such archives however presents interesting new challenges, such as how to organise this wealth of information, and make it accessible in meaningful ways. We present a Case Based Reasoning approach to creating a digital archive and adapting the representation of items in this archive. In doing so we are able to learn the important facets describing an item, and therefore improve the quality of recommendations made to users of the system. We evaluate this approach by constructing a user study, which was completed by industry experts and students. We also compare how users interact with both an offline physical case base, and the online digital case base. Evaluation of our representation adaptation, and our comparison of physical and digital archives, highlights key findings that can inform and strengthen the process for creating new case bases.

Keywords: Learning Refined Representation, Digitisation of Physical Archives, User Evaluation.

1 Introduction

The textiles industry is an important part of the local economy, history, and future of Scotland. Many prominent companies, manufacturing textiles for well-known designers, have existed for over 100 years. This heritage and experience is important, providing companies with a competitive edge over international rivals. However, while design processes and manufacturing technology have been kept up to date, the archiving of knowledge has suffered.

Typical archives throughout the textiles community are kept physically in storage rooms, similar to that in Figure 1. These physical archives can be difficult



Fig. 1. Typical Archive Room

to take full advantage of, unless someone knows exactly what they are looking for, and where it is. Sharing the knowledge held in separate archives can also be problematic, with many unaware of what may even be in an archive. There is wide interest across the community in how this knowledge can be made more accessible, and used as a source of inspiration for new textile designs.

In collaboration with Johnstons of Elgin, we have investigated how a physical archive can be transformed into a digital one. The first stage of this transformation is to understand the nature of the collection they have, and identify the important information available. Having obtained this information it is possible to create digital versions of physical assets, using photographs and descriptions. However, while this process makes information available, it does not organise it in a meaningful way. An index that highlights the important features of each digital asset, and facilitates the searching and browsing of assets, is required.

Case based reasoning (CBR) provides a structured way of modelling and learning from the past experience of users. Through capturing user behaviour and interactions, there is an opportunity to capture implicit knowledge about the content of the archive, that may improve the retrievals and recommendations. This implicit knowledge may be used to modify the asset index, and thus highlight the important features of each asset.

In this paper we discuss related work on the digitisation of physical archives, and how CBR has been used in retrieval and recommender systems. Our approach to understanding and selecting content for digitisation is then discussed. This approach includes two workshops with Johnstons of Elgin, a major textiles designer. We discuss our creation of the digital archive, and how cases can be defined to describe the selected assets. We then develop an implicit learning method that is able to refine the case representations based on user interactions. Next, we describe our user experiments, which evaluate our implicit learning method, and draw comparisons between the stakeholder and end-user

understanding of and interaction with the assets. We present the results of these user experiments, and finally draw some conclusions from this study.

2 Related Work

Although companies are only just realising the potential of digitally archiving their assets, several studies have investigated how this may be achieved. Evans [1] discusses the underlying requirements for the ‘perfect’ fashion archive, moving from physical to digital. A key finding of this work is that such archives do not necessarily need to contain all assets, but the organisation and interaction of key assets is important. Paterson [2] examines the success of digitising the House of Fraser fashion and textile archive. In this project, assets were added to a digital library, but catalog-book style indexes were relied upon to navigate the collection. The author concludes that for such an archive to be successful, more sophisticated indexing techniques and interactions are required. This view is further supported by Brown [3], who notes that sophisticated searching functionality is required for the successful creation of a digital archive.

CBR has previously been used successfully to provide searching functionality for e-commerce systems [4,5]. Such systems are somewhat similar to a textiles archive, however user motivation may differ. Case based reasoning has also been used to construct recommender systems, helping users to navigate a digital collection of items [6]. One main advantage of introducing CBR into a recommender system is that each item is no longer static in the collection. As users interact with the system, creating new problem queries and new solution recommendations, the system adapts to take advantage of this knowledge [7]. Such dynamic behaviour is critical to the successful creation of our digital archive.

Milne et al. [8] introduce this dynamic behaviour in an image retrieval system, by modifying the case representation weights after each successful query. Case weights that are consistently not aligned with associated queries, are diminished, thus removing noise from case representations and improving the overall system. Ontañón and Plaza [9] define a similarity measure based on the anti-unification of two structured cases being compared. This approach further highlights how shared information, in this example between two cases, may be used to further inform a CBR system.

3 Selection of Content for Archive

At the onset of this project no case base, or even digital versions of assets were available. The initial phase therefore was to engage with the owners of the data, explore, and formally define what each of the assets meant to them. To achieve this two formal workshops were conducted with the stakeholders: the first workshop to understand what an asset is, and how it may be described; the second workshop to understand how the stakeholders define relationships between assets. The results of these two workshops allowed us to create our initial digital case base.

3.1 Stakeholder Workshop 1: Asset Descriptions

The objective of the first user workshop was to identify what exactly assets are, and how they can be described. Staff from the company creating the archive were asked to contribute assets that they considered to be relevant to a digital archive. These contributions were initially in the form of one or more photographs, and a total of 150 items were contributed. From these 150 assets, justifications for each to be included in the archive were presented, and the 30 most interesting assets were selected to focus on by popular vote.

Example assets that were selected for inclusion in the archive range from textiles related, such as looms, tartan, and bale hook, to company related, such as history / heritage, and original skills and crafts. The images associated with each asset were clipped into plant pots, as illustrated in Figure 2, allowing everyone at the workshop to easily interact with and move around each of the items.



Fig. 2. Items Contributed by Stakeholders

The next step in this workshop was to gain an understanding of how each item may be described. With the assets clearly laid out to encourage discussion, a dialog with the stakeholders began to identify some key themes and commonalities amongst the assets. These themes were then translated into a set of labels, which the stakeholders used to categorise the collection. The 4 labels that emerged were future, luxury, heritage, and sentimental. The participants were then all given plant labels and post-it notes. The plant labels were used to annotate the items with key terms that small groups in the workshop discussed and defined, and the post-it notes allowed people to provide more detailed free-text descriptions. The outcome of this workshop formed the initial structure of our case representation for an item, consisting of the image, labels and tags, and free text.

3.2 Stakeholder Workshop 2: Asset Relationships

Having selected meaningful assets and obtained suitable descriptions, the second workshop aimed at understanding the structure of the collection as a whole. One important goal towards creating a useful archive is that it is intuitive and easy to navigate in a meaningful way. In CBR systems, this is most commonly achieved by introducing a meaningful semantic similarity measure.

To understand how the assets relate to each other, and gain insight into what similarity means, the items were again placed on the floor in their plant pots. The participants were split into their 4 respective departments; Retail, HR & Finance, Production, and Design. Each department was then asked to use coloured tape to form a map of how each item relates to another, illustrated in Figure 3. Each colour indicates how the stakeholders define relationships within three separate interest groups: future, heritage, and tourist. These groups were considered because they were believed to be the three main interest areas that end users could be classified into.

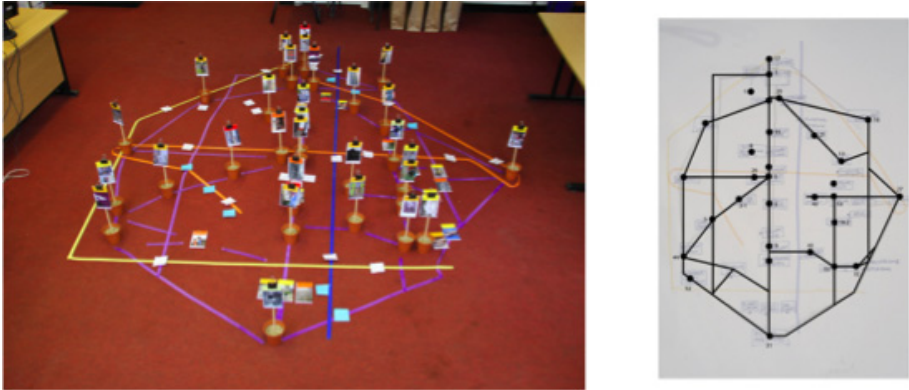


Fig. 3. Item map created by retail team (left) and transcription (right)

The construction of the maps was initiated by selecting start points from the 30 most popular assets. The edges were then added based on how the stakeholders believed each of the interest groups would navigate the archive, moving through related nodes. All of the assets were available for inclusion, but typically only between 30 and 40 were mapped. There was a general consensus amongst the 4 departments as to what assets were included. However, the way in which they were included differed between each department. For example, the production department produced a map which resembled a production line, while the retail team designed a map which was much more exploratory, similar to how someone may browse a shop. Although there were differing views across departments regarding which assets were relevant to each other, these maps provide interesting data which may be compared with usage patterns of the end-users.

4 Creating an Initial Case Base

The outcome of the workshop exercises was a definitive set of assets that are used to construct the initial case base. We represent this archive of assets as

a case base of cases. The final set of features that were used for each case are illustrated in Table 1. There are three types of features in our case structure; free text, tags, and images. Free-text features were processed by tokenising, stemming, and removing stop words from the data. These processed tokens were then used to construct a single term-frequency vector for each separate feature. The vectors for each feature were weighted using TF-IDF, and finally normalised.

Table 1. Case Representation

Type	Feature	Description	Vocabulary Size
Free Text	Title	The name of the item	144
Free Text	Description	Post-it note description	393
Free Text	Justification	Reason the item was chosen	344
Free Text	Other	Any other relevant descriptive information	72
Free Text	Facts	Interesting facts about the item	160
Tag	Aspects	Part of company the item relates to	14
Tag	Labels	Future, luxury, heritage, or sentimental	4
Tag	Terms	Key descriptive tags	59
Image	Main Image	SIFT Image features for main image	100

Tags were not tokenised or stemmed, and no stop words were removed. The reason for this is that the vocabulary sizes are much smaller, and each instance of the feature was typically only between one and three words long. Term-frequency vectors were again created for tags, and TF-IDF weighting and vector normalisation applied.

Images were indexed using the Scale Invariant Feature Transform (SIFT) algorithm [10], which detects and describes local features within an image. These local features were then clustered using the k -means algorithm, and a cluster-frequency vector created. The SIFT algorithm was chosen to describe images because it is well proven across many domains, and after clustering provides a representation structure which is similar to our textual representations.

The selection of assets from workshops, with free text, tags, and labels, together with the image provides the cases for the digital archive case base.

4.1 Content Similarity

We develop two methods for users to interact with the archive system: querying the items directly, and through a recommender system. Users provide a search query by typing search terms into a query box. These search terms must then be structured so that they may be used to access our case index. To achieve this, we construct a new temporary case from the query. In this case, each feature dimension matching a tokenised search term is incremented by 1, and all other dimensions are set to 0. For example, if the i^{th} dimension of the label feature

describes the frequency of term ‘tartan rug’, and the user queries for ‘red tartan rug’, then the term frequency is 2.

The recommender system uses a query-by-example approach, and thus the case describing the asset a user is currently viewing is used as the search query. To query our case base, using either a temporary or example case, we average the cosine similarities between each pair of individual feature vectors, calculated as:

$$\text{similarity}(Q, R) = \frac{\sum_{f=1}^F \frac{Q_f \cdot R_f}{|Q_f| \cdot |R_f|}}{F} \quad (1)$$

where Q is the query case, R is the potential retrieval, F is the number of features in a case, and Q_f and R_f are the f^{th} feature-vector of the query and retrieval cases.

5 Implicit Learning Method

With an archive case base constructed, features extracted, and similarity measure defined, the next stage is to refine methods used to navigate and interact with the archive. Although the workshop map could possibly be used as an initial refinement for similarity, the map covers only 20% of the assets used. To overcome this problem with missing edges, we propose a learning method which takes advantage of the implicit feedback created as each user interacts with the system.

5.1 Learning Feature Dimension Weights

The similarity measure proposed provides a good starting point to allow users to interact with the system. However, the system is simply using all of the information provided, whether it is relevant or not. The term frequencies of each feature dimension are based on the views of a relatively small number of participants at the workshops, and may be biased towards their opinions as employees. To overcome this problem we develop a new refinement method, which is able to learn implicitly which parts of a feature-vector are most important.

As each user interacts with the system, each query-retrieval pair they follow is stored. As each new user clicks on a retrieval, the feature-vectors of the retrieved case are refined. In the case of a recommendation, where the query is an existing case, the query case is also refined. This refinement is based on the information which is common to both query and recommendation, and how often the pair appears within the stored user interactions.

Let p denote the number of times a query-retrieval pair has previously been successful, and t denote the number of times a query has been successful. The strength s of refinement is calculated as

$$s = 1 + \log \left(\frac{p}{t} + 1 \right) \quad (2)$$

which is in the range 1 to 1.69. Refinement is then applied to information which is common to both the query and retrieval as

$$Q_{fi} = \begin{cases} s \cdot Q_{fi} & \text{if } Q_{fi} > 0 \text{ and } R_{fi} > 0 \\ Q_{fi} & \text{otherwise} \end{cases} \quad (3)$$

and

$$R_{fi} = \begin{cases} s \cdot R_{fi} & \text{if } Q_{fi} > 0 \text{ and } R_{fi} > 0 \\ R_{fi} & \text{otherwise} \end{cases} \quad (4)$$

where f is the feature-vector being refined, and i is the i^{th} dimension of feature vector f . The conditions ' $Q_{fi} > 0$ and $R_{fi} > 0$ ', and ' $R_{fi} > 0$ and $Q_{fi} > 0$ ' assure that only feature dimensions which are shared are refined. After the weights are modified, the vectors are re-normalised. This refinement process is illustrated for a single feature-vector in Figure 4.

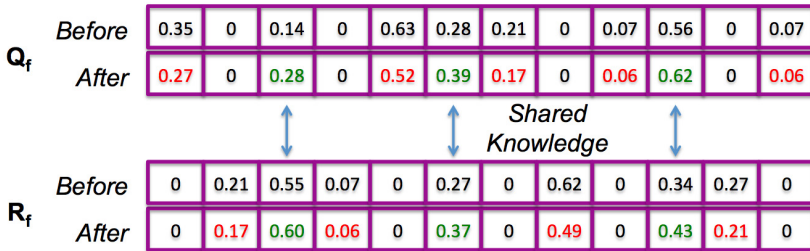


Fig. 4. Implicit Learning of Feature Dimension Weights

Normalising each feature-vector after increasing the weights on shared knowledge, means that the weights of unshared knowledge will be decreased. The amount of information in the feature-vector remains constant, but is moved from one dimension to another. The effect of increasing the weights of shared dimensions in a feature vector will mean that the query-retrieval pair will become more similar to each other. However, this modification may also have further effects throughout the entire search space, pushing both query and retrieval to be more similar to some unknown items, and less similar to other items. This behaviour is desirable, since it is a consequence of refining an item's representation to more accurately reflect how it relates to the collection.

6 User Experiments

To evaluate the methods discussed, an archive website was developed and made available to selected users online. The goal of this website was to measure and evaluate several objectives. Firstly, we wish to evaluate the implicit learning method that has been developed. However, we are also interested in the user engagement with the archive system. In this section the website is described, and the evaluation method for our user trial is presented.

6.1 User Engagement with the Archive Website

The system that was developed allowed users to navigate the archive in several different ways, shown in Figure 5. A search box was provided in the top right corner to allow free-text searching of the archive. As a user views an item in the archive they are also provided with a set of recommended items, below the item they are currently viewing. In Figure 5, a user is viewing the ‘Dye Pot’ item, and this item is used as a query for the recommendation list. These recommendations were generated using the similarity measure defined in Equation (1), and as users followed recommendations the implicit learning defined in Equation (4) was applied.

A recent trend in many catalog-based websites is to allow users to bookmark, or ‘favourite’ items that they want to return to. To gain a fuller insight into how users decide to interact with our archive system therefore, we also implement this feature. As users browse the collection, they can add any item to a favourites bar which is accessible from every page.

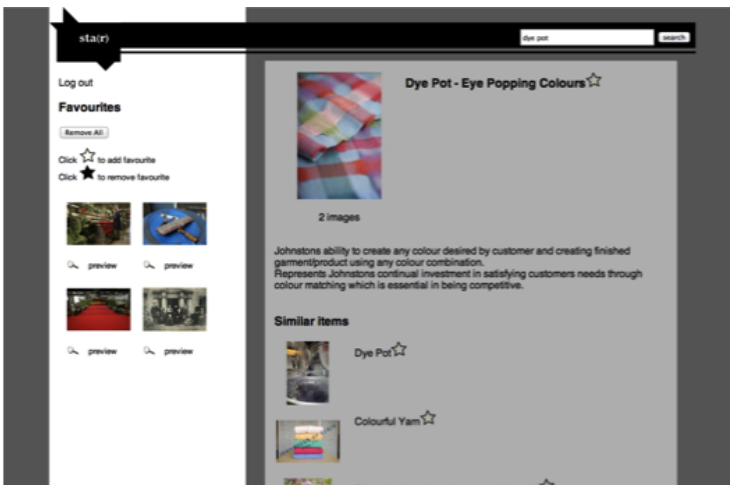


Fig. 5. Smart Textiles Archive System

6.2 Recommendation Quality

To evaluate the effect of our implicit learning method we measure how effective the recommender system is. Ideally, we would hope that as the items become more refined, the most relevant recommendations will appear closer to the top of the ranked recommendation list. To measure this we therefore calculate the average recommendation rank of a single query item as

$$\text{average recommendation rank} = \frac{\sum_{j=1}^J \text{rank}(j)}{J} \quad (5)$$

where j is an instance of the query item being used, J is the number of times the query has ever been used, and $\text{rank}(j)$ is the position that the recommendation a user clicked on was presented. We report the mean average recommendation rank across all queries, obtained at varying levels of refinement.

7 Results: Effects of Implicit Learning

The archive was made available online for 1 month, and invitations were sent to both industry experts and students from the textiles field. Over this trial period, 8 industry experts from 5 separate companies, and 11 students from 2 universities participated in the study. Each user was asked to complete several investigative tasks, for example, finding out about a certain type of material using the archive.

Figure 6 shows the effect that our implicit learning method has on recommendation rank. The vertical axis shows the average position of the recommendation that was clicked on by a user. The horizontal axis shows the number of times that the query item has been modified by the implicit learning method.

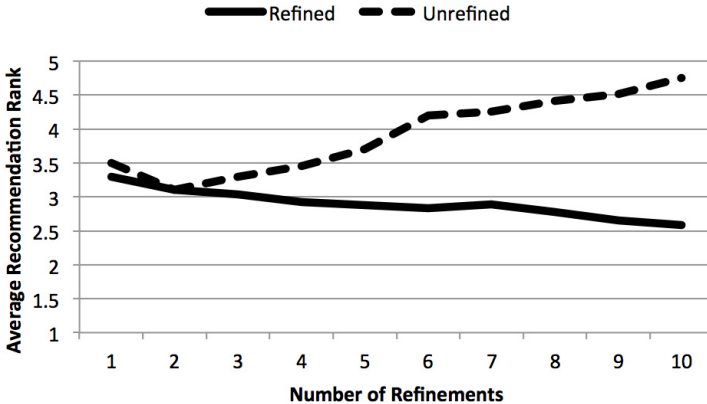


Fig. 6. Recommendation Rank After Implicit Learning

The solid black line in Figure 6 shows the average recommendation rank achieved after each refinement using our implicit learning method. After a single refinement to the query, the average recommendation rank is 3.3, meaning that on average, a user will click on either the 3rd or 4th recommendation in our ordered list. As the system is used more, and more refinements are made, implicit learning has the effect of lowering this average rank. After 10 refinements the average recommendation position the user clicks on is 2.6.

The dashed line in Figure 6 shows the ranking that recommendations would have if the refined representation is not used. This line is not flat as expected, because all clicks were logged for recommendations made using our refined representation. The difference between the solid and dashed line does however illustrate the power of our implicit learning method. After 10 refinements the average rank of a good recommendation is reduced by 1 position in the ranked list. This makes the browsing experience for the user easier, and helps them to find the items they are interested in more quickly.

Reducing the average rank of a recommendation is not only because the query has been refined; we apply our implicit learning method to both the query and recommendation. This is an important contributing factor to the results observed in Figure 6. When a user is viewing an item, and follows a recommendation, the recommended item is refined. This item then becomes the query for the recommendations provided on the page which is loaded, and the more meaningful refined representation can be used. This illustrates the power of refining a representation, compared to simply re-ordering results based on previous cases.

8 Results: User Engagement with Archive

Further to evaluating our implicit learning method, we are also interested in how users engage with the digital archive. In our workshops to establish the initial case base, the stakeholders categorised each item as primary, secondary, and supplementary. The stakeholders also constructed a map of items, based on the relationships that they considered to be relevant to someone browsing the archive. These workshops provide a wealth of information that can be compared to how the industry expert and students engaged with the digital archive.

8.1 Physical and Digital Relationship Maps

Using the user behaviour logged by our digital archive, we are able to construct the recommendation map illustrated in Figure 7. Each node in the map represents an item, and each edge represents a followed recommendation. The direction of the arrows represents the direction query to recommendation, where the arrow points to the recommended item. The position of each node in the map is determined using multi-dimensional scaling, where each item constitutes a new dimension, and the distance between items is inversely proportional to the number of times the query-recommendation pair occurred.

In the recommendation map three distinct clusters can be observed, each annotated in Figure 7. Perhaps unsurprisingly these clusters emerge as some of the major themes discussed throughout the workshop phase. The company is a textiles manufacturer, and the importance of history and heritage was agreed upon by all stakeholders. The design / process cluster occurs as a result of the distinct nature of the production and design teams' contributions.

The brightness of each node in the map indicates how frequently the item was viewed, where lightest is most frequent. This helps to identify hubs within

the items, which are used frequently as both a query and a recommendation. This shading also helps to identify the items that may be considered primary, secondary, and supplementary, in a similar manner to the offline workshops.



Fig. 7. Map of Followed Recommendations

Figure 8 illustrates the map created by the stakeholders during the workshop. The position of each node in the map is frozen to match the positions defined in Figure 7, facilitating comparison of the offline and online item relationships. One thing that stands out instantly between the two maps is that the stakeholder map contains many fewer edges. The reason for this was primarily due to time constraints, and the stakeholders focussed on the primary assets.

The length of edges in the stakeholder map are in general very long. This indicates that the relationships of items as perceived by the stakeholder does not match the relationship engaged with by the online users. The history / heritage cluster is well understood by the stakeholder, illustrated by the shorter edges, but there is confusion between the design and textiles clusters. Within the company, the realisation that product and design process should be considered as separate areas of interest was not made. From a business perspective these two areas are very closely connected, but to online users they are not. Finally, it can be observed that some nodes have no edges in the online map. This illustrates a further misunderstanding from the stakeholder perspective of how an end user would want to traverse the digital archive.

These observations of differences between the stakeholder workshop and online usage further illustrate the importance of our representation refinement method. At the point of putting the system online, each item begins with an unrefined representation, that has been constructed by the item owners. While this representation correctly describes an item, it does not necessarily reflect a description

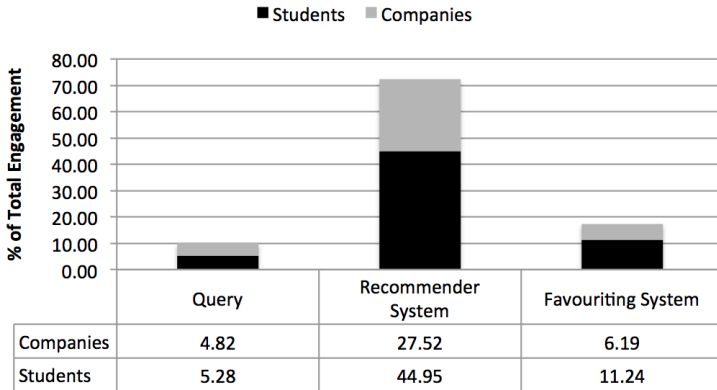


Fig. 9. Modes of User Engagement

looking for. Feedback collected from users indicated that they were using the archive as a form of inspiration: starting with an item they recognise found through the query system, and then exploring the similar items suggested by the recommender system. This places the recommender system as a key component of the archive, and as such implicit learning is essential to take advantage of the end user interests to improve the system.

9 Conclusions

We have conducted a study of how a physical archive can be transformed into a digital archive. This study includes how stakeholders of the archive understand their knowledge, and how this knowledge may be transformed into a meaningful digital system. Through the workshop phase of the project, important insights and knowledge about the physical collection was learned, which facilitated the construction of the archive.

However, results show that knowledge learned at the workshop phase is biased towards the stakeholder perception of the items and their relationships. We have developed a new implicit learning method that can refine the knowledge learned from the workshop phase, and adapt the representations of an item to reflect end user interests. Our method strengthens the shared knowledge between queries and recommendations, thus strengthening the knowledge that is important to their relationship. Over time this process refines the entire collection so that the representations are weighted appropriately, and a more meaningful similarity measure can be calculated. The result of this refinement process is that relevant items appear closer to the top of ranked recommendation lists, thus enabling the end user to find interesting relevant items more quickly.

Investigations of user engagement with our archive highlights the importance of a recommender system to exploratory and inspirational systems. Quite often the user will have a vague idea of what they are looking for, but do not know how

to describe it as a well defined query. This conflicted with the engagement patterns predicted by stakeholders, who knew their data well enough that perhaps a query system may have worked well internally. However, our study further highlights the need for a retrieval system to learn what is important to the end-user base, managed successfully by our implicit learning method.

Acknowledgements. The authors wish to acknowledge AHRC for its support through its Digital Transformations Research Development award for this project (AH/J013218/1) and the industrial collaborator Johnstons of Elgin for their contribution of data and stakeholder workshops.

References

1. Evans, C.K.: Developing the perfect fashion archive. In: 1st Global Conference: Fashion Exploring Critical Issues (2009)
2. Paterson, C.: Selling fashion: realizing the research potential of the house of fraser archive, university of glasgow archive services. *Textile History* 40(2), 170–184 (2009)
3. Brown, C.: Digitisation projects at the university of dundee archive services. *Program: Electronic Library and Information Systems* 40(2), 168–177 (2006)
4. Nisanbayev, Y., Ko, I., Abdullaev, S., Na, H., Lim, D.: E-commerce applications of the hybrid reasoning method. In: *International Conference on New Trends in Information and Service Science, NISS 2009*, pp. 797–801. IEEE (2009)
5. Vollrath, I., Wilke, W., Bergmann, R.: Case-based reasoning support for online catalog sales. *IEEE Internet Computing* 2(4), 47–54 (1998)
6. Bridge, D., Göker, M., McGinty, L., Smyth, B.: Case-based recommender systems. *The Knowledge Engineering Review* 20(03), 315–320 (2005)
7. Lorenzi, F., Ricci, F.: Case-based recommender systems: A unifying view. In: Mobasher, B., Anand, S.S. (eds.) *ITWP 2003. LNCS (LNAI)*, vol. 3169, pp. 89–113. Springer, Heidelberg (2005)
8. Milne, P., Wiratunga, N., Lothian, R., Song, D.: Reuse of search experience for resource transformation. In: *Workshop Proceedings of the 8th International Conference on Case-Based Reasoning*, pp. 45–54 (2009)
9. Ontañón, S., Plaza, E.: On similarity measures based on a refinement lattice. In: McGinty, L., Wilson, D.C. (eds.) *ICCB 2009. LNCS*, vol. 5650, pp. 240–255. Springer, Heidelberg (2009)
10. Lowe, D.G.: Object recognition from local scale-invariant features. In: *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157. IEEE (1999)