Summary

As audio-visual archives become accessible to the wider internet audience, successful retrieval of archive items is becoming an increasingly challenging task. This is mainly caused by the lack of adequate annotations. Users cannot find what they are looking for because annotations are either not present or too expert-centric. Video tagging games are an attempt to alleviate this problem. Engaging the internet community to tag the videos addresses the issue of scarcity of annotations and yields tags that are community-centric. The focus of this thesis is on the role of game tags in audio-visual archives and their added value for video retrieval.

In the typical audio-visual archive ecosystem the game tags will coexist with professionally curated annotations. To understand what game tags bring to the table as opposed to the catalogers’ annotations we study the differences between them along two dimensions. The first dimension is the terminology: we perform a quantitative study to estimate the terminological gap between taggers and the catalogers. The second dimension is the descriptive scope i.e. which aspects of the video content are described and at what level of abstraction and granularity. We carry out a qualitative study in which we analyse the descriptive scope of the game tags. Our analysis shows that there is a terminological gap between the players of Waisda? and the professionals. This makes the game tags a valuable asset as they are from the users and for the users and can improve the access to the videos in the case when the ones accessing are the users themselves. Furthermore, we find that game tags predominately describe instances (objects) in the video and rarely scenes. This result provides important intuition about the usefulness of the game tags for video retrieval which is the focus of the remainder of the thesis.

Throughout this thesis we address two prominent search scenarios that arise in practice. The first scenario is retrieving video fragments that feature visual appearances of objects. In fact, this is very much like the instance search task from TRECVID with the difference that query is formulated in text and not by visual example as in TRECVID. The second scenario we
consider is *topical search* i.e. retrieval of video fragments that are about a given topic.

In Chapter 3 we study how effective are the game tags in addressing the first scenario. To this end, we design Cranfield-style experiments which require collection of “documents” (in this case video fragments tagged by users via *Waisda*?), set of queries which we derived from real-life query logs, and relevance judgments indicating which “documents” in the collection should be returned for each query. The results from the evaluation show that game tags are extremely well-suited for retrieving fragments that visually depict a given object of interest. They outperform the other types of metadata (including professional annotations) in pairwise comparison. We also establish that there is a difference in the focus and the scope of the game tags compared to the professional annotations; game tags being more fine-grained and professional annotations being more coarse-grained referring to the entire video and describing the prevalent topics. This property of the game tags makes them particularly appealing for business use-cases such as discovering/selling stock footage.

In Chapter 4 we evaluate the performance of game tags for topical search using another set of Cranfield-style experiments. The results demonstrate that raw, unprocessed game tags are not well suited for retrieving video fragments based on topic. While the search recall is satisfactory, the search precision leaves much to be desired. This is mainly caused by the presence of game tags which are not valid topical annotations. The fast pace of *Waisda*? usually results in players contributing mainly non-topical tags. To combat this effect in the topical search scenario, we characterize the quality of user tags as topical annotations and identify several tag features (filters) that serve as indicators whether game tags are useful for retrieval. Our results show that after filtering, game tags can emulate the retrieval performance of a baseline system that utilizes manually crafted metadata for search. An important consequence of this finding is that tagging games can provide a cost-effective alternative in situations when manual annotation by professionals is too costly.

The general conclusion of this thesis is that the game tags improve retrieval and on average perform better than the manually-crafted metadata and closed captions. This statement comes with three caveats.

First, in our setting the taggers and the searchers originate from the same population which results in significant terminological overlap between the keywords used for search and the tags ascribed to the videos. We believe this to be a strong contributing factor for the success of the game tags for search in the context of our study. While this may look like a limitation at first, in practice it is often not, because the terminology employed when
tagging is typically sufficient to cater to the needs of all but the expert searchers which may use more specialized (specific) terms when searching.

Second, in this work we considered only keyword-based search thus the said improvement in retrieval is limited to this modality. At the time of writing, other search modalities such as content-based video retrieval are gaining more attention, however keyword-based search is still the most prominent one, especially on the Web.

Third, in this work we focussed on single domain (genre), non-fictional TV Comedy. Consequently, the obtained results are bound to this domain. However, the qualitative analysis of the tags revealed that they are usually general concepts describing visual objects and as such are not so tightly coupled with the genre of the video content they are describing. Our expectation is that the game tags will perform well for most of the non-fictional TV genres which appeal to a wider audience. It remains to be seen whether this will be the case for the fictional genres where more symbolism and abstract expressions can be expected. We hypothesise that without properly instructing the taggers or targeting a niche sub-community of experts the outcome will be less satisfactory.

The fact that game tags can successfully support video retrieval is of particular importance for the audio-visual cultural heritage domain. The national audio-visual archives throughout Europe have hundreds of thousands hours of digitized audio-visual material that is not readily accessible because of lack of annotations. The audio-visual archives can now launch Waisda?-like crowdsourcing campaigns confident that the collected tags will enable access to their collection items.