MIRACLE's working group
"Child Protection in Social Media"

- Insights from the kick-off workshop, 9th Sep 2015, Berlin -

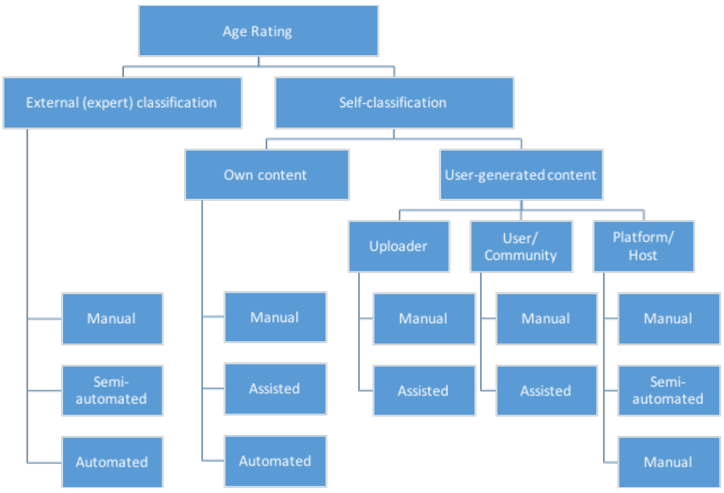
Background

Platforms for user generated content (UGC) play a major role for children and young people nowadays. In such environments the service providers usually do not offer any inappropriate content themselves. Instead, they only provide the framework or platform for relevant content, uploaded by the individual users. And yet many UGC services have implemented tools for the protection of younger users on their platforms, such as user-generated content classification and reporting tools for inappropriate content.

In the context of the MIRACLE project (www.miracle-label.eu), a technical pilot project developing and implementing interoperable electronic age classifications, the working group "Child protection in Social Media" has recently been established. The working group's objective is to collect the UGC providers' current approaches in the area of age classifications and labelling, to analyse experiences and outcomes of these approaches and to identify best practice solutions. The working group acknowledges that there is no technical one-size-fits-all solution due to different innovations, international differences in operating procedures and (often internal) standards, different evolutionary levels of implementation as well as their contexts and options. We start from the perspective that the existing approaches rather need to be strengthened by furthering their potentials through technical interoperability. By standardizing age classification-related meta data of user generated content, the working group aims at improving consumer information and child protection tools on web 2.0 platforms.

Against this background the participants of the kick-off workshop convened on 9th September 2015 to discuss current classification and labelling practices on UGC platforms as well as the potentials of using interoperable classifications.

Diagram: Categorisation of content classification



Key Takeaways of the workshop

There's a lot of classification knowledge stored by user-generated content platforms, but it's usually locked away.

- All major UGC platforms use rating or flagging tools for inappropriate, offensive, illegal content. This results in the production of classification information on the side of the platform provider. Hence, there are massive amounts of knowledge regarding the uploaded content stored at UGC platforms. This knowledge is usually used internally only, leading to siloed classification knowledge.
- Most platforms use this knowledge to classify content in two or three categories: Appropriate for all age groups, appropriate for some age groups, not appropriate at all. Regional differences (e.g. when it comes to taste or decency) or reasons for the tags/flags are usually lost when UGC providers make use of the reporting tools in these ways.

There are relationships between the steps of content classification and content labelling, but usually any form of content classification can use any form of electronic labelling.

- The step of letting users classify content is independent from the act of labelling, in general. UGC providers can opt for almost any form of classification procedure and then provide this information by a labelling technique of their choice.
- An aspect where classification and labelling are entangled is quality and validity of the age classification information. If the data basis is of high quality and validity, so is the label based on this information. If the data basis is weak or the result of (partial) misuse, this might spread over to the information contained within a label, depending on the classification approach taken.

Current user-generated classification approaches rely on ex-ante uploader self-classification or ex post community classification.

- In practice the majority of UGC platforms uses tagging and flagging tools as reporting mechanisms for inappropriate content. In such cases, the providers rely on community-based classification after the content has been made available by an uploader. Some platforms enable the uploaders to classify the uploaded content by themselves (e.g. not suited for children). Few platforms make additional use of algorithm-based classification.
- Some UGC platforms offer both unrated user-generated content and professional media content that has been rated in advance by an external classification body. These platform providers have to process both external and community-based classification information.

Current practices of using the user-generated classification knowledge is limited to internal measures. Public labelling is non-existent.

- The (internal) processing of user-based content classification is mostly three-fold:
  - Illegal and inappropriate content or uploads infringing with the Terms of Service will usually be deleted.

    - Content that is deemed unsuitable for younger age groups get age-gated, e.g. by making it available to registered users only, by applying a warning sign or extra layer against accidental display, or by delisting it from search results for users who have activated some form of safe search feature.
    - Content that is not deemed inappropriate or in breach of the platforms content guidelines will stay online as is.
    - (Some platforms actively flag content that is directed at children, making such positive content retrievable within specific children sections or apps.)
- The discussion during the workshop showed that the user-generated classification knowledge of the platforms regarding the content that stays online (be it in an age-gated way or not) usually is lost for external demanders and consumers. Specifically, the knowledge about potential reasons for a problematic content won't be transferred any longer (even if they are still existent internally on side of the platform provider).
- Opening this knowledge to external demanders, however, might suit the individual user's needs much more than one-size-fits-all internal blocking/gating approaches. Interoperable age classification information can provide the opportunity to maximise the value and benefits of user-based age ratings.

MIRACLE's added value for social media and user-generated content platform providers

Processing, transferring and exchanging age classification data in one single format

MIRACLE enables companies that classify their content themselves or with the help of their user communities to use a data specification that is highly adaptable to the specific needs. Using MIRACLE datasets ensures that all age classifications follow the same data scheme, making it easy to process, transfer and exchange classification information within a company, with external partners, or both.

Mapping of existing classification data without touching the underlying scheme

If a platform provider doesn't want to change his existing internal classification scheme or IT infrastructure, MIRACLE enables such companies to simply map the categorization to the MIRACLE specification without altering the underlying structures. This way, classification workflows and technical environments can stay the way they are, making MIRACLE an additional/alternative channel to provide age rating data.

Provision of classification data in a comprehensible, machine-readable format

By providing or transferring age classification information in a MIRACLE-based format, additional documentation and transaction costs due to proprietary formats are unnecessary. The more nodes of a network use MIRACLE, the less additional APIs or mappings have to be produced.

Streamlined Age Rating Storage and Processing

No matter how many different types of content a company provides, and regardless of the markets it is serving: By using MIRACLE-based age classification data, all the different age rating procedures used will always result in the same data format – without losing the original meaning and information of the underlying rating schemes.

Mapping and aggregating ratings from different sources and national schemes

If a platform provider handles media assets that have been classified by a variety of different regional or national legal frameworks, the different age ratings can become impossible to store and process in an efficient way; even more so in cases where external classifications come upon internal user-based age classifications. By using MIRACLE, content providers are able to map the different schemes on one vocabulary and aggregate the data in one dataset. For media asset management software this is an opportunity with a significant added value: One file, containing all the different age classifications, and machine-readable on basis of only one common vocabulary.

One data format for backend and frontend

The data platform providers make use of within their backend systems offer great opportunities for frontend needs and features, too (see below). In such cases the data format doesn't change, resulting in one data model fitting all needs – be it internally or demand-side driven.

Frontend labels for provided content - regardless of the platform

Providing the classification information to third parties and end users can be a great way of sharing classification expertise that would otherwise stay locked away, enabling the end-user to use third-party software or plugins to process the provided classification data locally. Implementing MIRACLE as a data format for electronic labels enables content providers to provide classification information in an interoperable format, enabling third parties to use the data without much hassle. Browser plugins, parental control software, embedding partners, intermediaries – they all make use of interoperable labels provided by the content providers.

Region-specific playout of age classifications

In many countries content providers are legally or informally obliged to classify and label their relevant content. For internationally active platform providers this results in the necessity to provide age labels depending on the user's country of residence. With aggregated, region-country-specific MIRACLE datasets such companies are able to provide exactly that age classification information which is relevant for the individual user.

API-based provision of classification information for third party software and external demanders

In cases of user generated content platforms, companies offer their users a large range of different content, resulting in highly different age classifications. In such situations, MIRACLE offers ways to provide MIRACLE datasets via an API endpoint, making the data specification both extremely slim and versatile. Demanders for API-based data provisions might be third party software or parental control software providers. The main advantage of using an API in these cases are

(a) slim datasets, since it only carries information regarding a specified content and

(b) traffic-saving, since the API will only be queried by demanders who request such information. Non-users of electronic labels won't experience any difference in content provision.

Next steps

- The working group plans to approach industry players to learn more about the strategies and hesitations of the industry regarding using UGC ratings, also and especially externally.
- It has been agreed that the working group will need a partner who will pilot a MIRACLE implementation in the context of a UGC platform. A potential candidate might be yourateit.eu (tbd).
- The working group will have to provide more information on the technical side of a MIRACLE-based implementation of electronic age labels.
- The working group will approach the ICT Coalition to discuss potentials and probable issues when it comes to introducing these endeavours to industry players on EU level.