

Commons und die Metadaten

Kategorienchaos statt auswertbare Metadaten

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt
- willkürliche Sprachwahl bei Benennung der Kategorien

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt
- willkürliche Sprachwahl bei Benennung der Kategorien
- Personenkategorien mit nicht-lateinischen Buchstaben besonders schwierig: mal in Heimatsprache, mal in englischer Transkription, mal in willkürlicher Transkription je nach Herkunft des Erstellers

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt
- willkürliche Sprachwahl bei Benennung der Kategorien
- Personenkategorien mit nicht-lateinischen Buchstaben besonders schwierig: mal in Heimatsprache, mal in englischer Transkription, mal in willkürlicher Transkription je nach Herkunft des Erstellers
- gerade zu Sportlern oft mehrere Kategorien unterschiedlicher Schreibweise vorhanden

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt
- willkürliche Sprachwahl bei Benennung der Kategorien
- Personenkategorien mit nicht-lateinischen Buchstaben besonders schwierig: mal in Heimatsprache, mal in englischer Transkription, mal in willkürlicher Transkription je nach Herkunft des Erstellers
- gerade zu Sportlern oft mehrere Kategorien unterschiedlicher Schreibweise vorhanden
- manche Kategorien völlig überfüllt, manche kleinteilig zersplittert

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt
- willkürliche Sprachwahl bei Benennung der Kategorien
- Personenkategorien mit nicht-lateinischen Buchstaben besonders schwierig: mal in Heimatsprache, mal in englischer Transkription, mal in willkürlicher Transkription je nach Herkunft des Erstellers
- gerade zu Sportlern oft mehrere Kategorien unterschiedlicher Schreibweise vorhanden
- manche Kategorien völlig überfüllt, manche kleinteilig zersplittert
- Suchen in mehreren Kategorien nur über externe Werkzeuge wie PetScan, keine solche Funktion in der Software integriert

Commons und die Metadaten



Nicht nur für Beitragende, auch für Nutzer großes Problem: Fotos finden!

- Kategoriensystem absolut nicht konsistent, wird nur punktuell benutzt
- willkürliche Sprachwahl bei Benennung der Kategorien
- Personenkategorien mit nicht-lateinischen Buchstaben besonders schwierig: mal in Heimatsprache, mal in englischer Transkription, mal in willkürlicher Transkription je nach Herkunft des Erstellers
- gerade zu Sportlern oft mehrere Kategorien unterschiedlicher Schreibweise vorhanden
- manche Kategorien völlig überfüllt, manche kleinteilig zersplittert
- Suchen in mehreren Kategorien nur über externe Werkzeuge wie PetScan, keine solche Funktion in der Software integriert
→ **kein Normalnutzer findet PetScan und kann es auf Anhieb bedienen**

Commons und die Metadaten



völlig überflüssige Kategorien:

- Category:September 2011 Germany photographs
- Category:Germany photographs taken on 2016-01-17
- Category:CC-BY-SA-3.0-DE
- Category:Media with locations

Commons und die Metadaten



völlig überflüssige Kategorien:

- Category:September 2011 Germany photographs
- Category:Germany photographs taken on 2016-01-17
- Category:CC-BY-SA-3.0-DE
- Category:Media with locations

Alle diese Informationen sind gespeichert und damit abfragbar, ohne Kategorien anlegen und zuordnen zu müssen.

Commons und die Metadaten



völlig überflüssige Kategorien:

- Category:September 2011 Germany photographs
- Category:Germany photographs taken on 2016-01-17
- Category:CC-BY-SA-3.0-DE
- Category:Media with locations

Alle diese Informationen sind gespeichert und damit abfragbar, ohne Kategorien anlegen und zuordnen zu müssen.

generell: Wenn ein Bot eine Kategorie ohne menschliches Zutun zuordnen kann, ist diese redundant, da es eine Suchfunktion dann auch könnte.

Commons und die Metadaten



Exkurs: Datenbank-Normalform

- 1. Normalform: Jedes Attribut der Relation muss einen atomaren Wertebereich haben, und die Relation muss frei von Wiederholungsgruppen sein.

Exkurs: Datenbank-Normalform

- 1. Normalform: Jedes Attribut der Relation muss einen atomaren Wertebereich haben, und die Relation muss frei von Wiederholungsgruppen sein.
- Beispiel: Die Adresse darf nicht als einzelnes Attribut verwendet werden, sondern muss in PLZ, Ort, Straße, Hausnummer etc. aufgeteilt werden.

Exkurs: Datenbank-Normalform

- 1. Normalform: Jedes Attribut der Relation muss einen atomaren Wertebereich haben, und die Relation muss frei von Wiederholungsgruppen sein.
- Beispiel: Die Adresse darf nicht als einzelnes Attribut verwendet werden, sondern muss in PLZ, Ort, Straße, Hausnummer etc. aufgeteilt werden.
- bezogen auf Commons: die Kategorie „Female biathletes from Germany“ dürfte nicht existieren

Exkurs: Datenbank-Normalform

- 1. Normalform: Jedes Attribut der Relation muss einen atomaren Wertebereich haben, und die Relation muss frei von Wiederholungsgruppen sein.
- Beispiel: Die Adresse darf nicht als einzelnes Attribut verwendet werden, sondern muss in PLZ, Ort, Straße, Hausnummer etc. aufgeteilt werden.
- bezogen auf Commons: die Kategorie „Female biathletes from Germany“ dürfte nicht existieren
- Richtig wären die einzelnen Kategorien „Female“, „Biathlete“, „Germany“

Commons und die Metadaten



Sehr viele Fotos auf Commons enthalten inzwischen wertvolle Informationen in den Exif- und IPTC-Daten:

- Kameramodell, Objektiv, Belichtungseinstellungen
- Geokoordinaten, menschenlesbare Ortsangaben, Aufnahmezeitpunkt
- Subjectcodes, Scenecodes, Genre codes
- Stichwörter, Bildtitel, Untertitel

Commons und die Metadaten



Sehr viele Fotos auf Commons enthalten inzwischen wertvolle Informationen in den Exif- und IPTC-Daten:

- Kameramodell, Objektiv, Belichtungseinstellungen
- Geokoordinaten, menschenlesbare Ortsangaben, Aufnahmezeitpunkt
- Subjectcodes, Scenecodes, Genre codes
- Stichwörter, Bildtitel, Untertitel

Alle diese sinnvollen Daten können nicht genutzt werden, da sie noch immer nicht suchbar / filterbar sind!

Commons und die Metadaten



Bei einer Auswertbarkeit der Exif- und IPTC-Daten ließen sich komplexe Abfragen generieren, z. B.

- Basketballfotos aus dem Jahr 2016 aus Deutschland
- Konzertfotos mit ausgelöstem Blitz aus der Schweiz
- Fotos von Orchideen mit Offenblende und Brennweite <50 mm
- Fotos mit einer EOS 1DX und Sigma 120-300 mm Objektiv
- Fotos der deutschen Handball-Frauen bei Spielen der Bundesliga, im DHB-Pokal oder der Champions League

Commons und die Metadaten



Und wie könnte Wikidata helfen?

Commons und die Metadaten



~~Und wie könnte Wikidata helfen?~~

Wann wird Wikidata helfen?