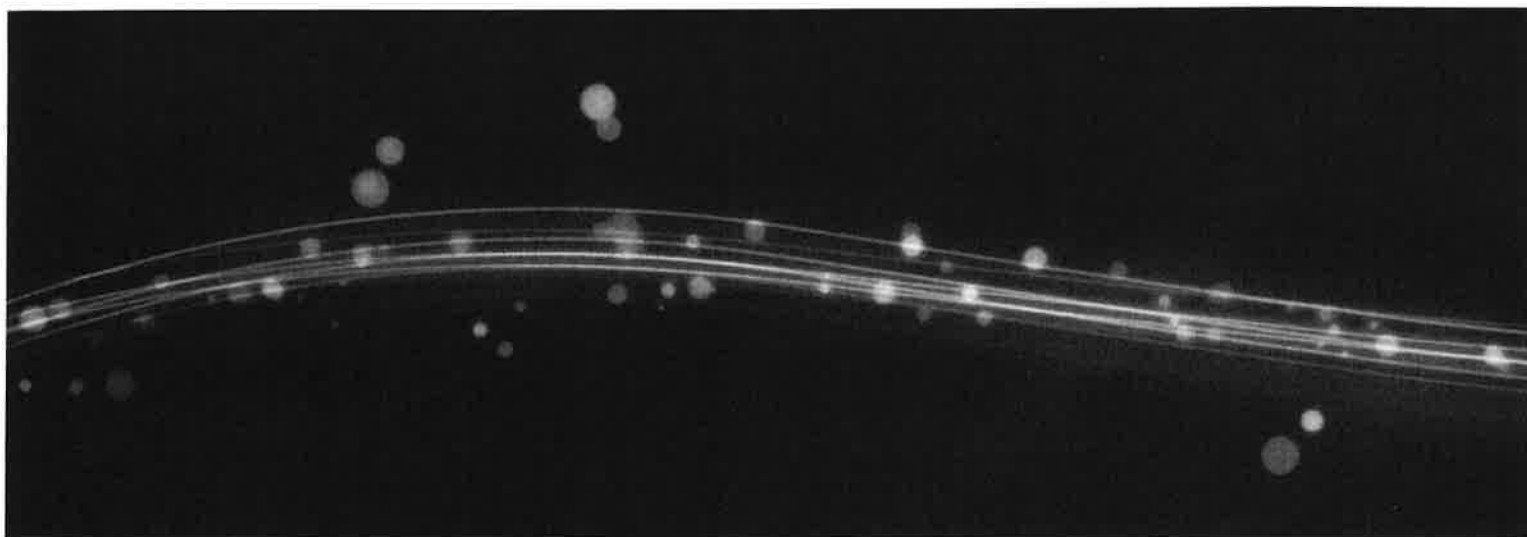


Angebot

IT-Tool zur Recherche potenziell rechtswidriger Inhalte im Internet – Überarbeitetes Angebot

Teil A: Konzeption, Projektvorgehen, Leistungen, Vergütung

Für die
Landesanstalt für Medien NRW
Zollhof 2
40221 Düsseldorf



Angebot

IT-Tool zur Recherche potenziell rechtswidriger Inhalte im Internet – Überarbeitetes Angebot

Teil A: Konzeption, Projektvorgehen, Leistungen, Vergütung

Auftraggeber: Landesanstalt für Medien NRW
Straße: Zollhof 2
PLZ/Ort: 40221 Düsseldorf

Erstellt durch: Condat AG
Alt-Moabit 91d
10559 Berlin

Vorstand: Lars Fahrenholz, Dominik Zimmermann
Aufsichtsratsvorsitzender: Peter Herrmann
HRB Berlin Charlottenburg 83319

Telefon: 030 3949 1270
Fax: 030 3949 1300

Angebotsdatum: 11.06.2020

Erstellt: 11.06.2020

Inhaltsverzeichnis

1	Ausgangssituation und Projektbeschreibung	5
1.1	Der Auftraggeber	5
1.2	Gegenstand des Angebotes	5
1.3	Die Änderungen im überarbeiteten Angebot	6
2	Lösung und Leistung	9
2.1	Der Lösungsansatz im Überblick	9
2.2	Recherche	10
2.2.1	Automatisierte, periodische Internet-Suchen	10
2.2.2	Spezialisierte Suchen in sozialen Netzwerken	11
2.2.3	Crawling individueller Quellen	12
2.2.4	Multimediale Quellen	12
2.2.5	Black- und Whitelisting von Anbietern und Quellen	13
2.3	Analyse und Identifikation der Inhalte	13
2.3.1	Wissensmodell	13
2.3.2	Semantische Analyse	14
2.3.3	Klassifikation und Identifikation	15
2.3.4	Kontinuierliche Verbesserung - das lernende System	16
2.3.5	Integration externer Dienste	17
2.3.6	Abschließende Bewertung und Kategorisierung	19
2.4	Bearbeitung und Verwaltung der Inhalte (Datenbank)	19
2.4.1	Datenbank der identifizierten Inhalte	19
2.4.2	Workflow zur Abarbeitung der identifizierten Inhalte	21
2.4.3	Benutzerverwaltung	24
2.5	Architektur	25
2.5.1	Recherche-Komponente	25

2.5.2	Analyse & Identifikation	26
	Vorgehensvorschlag und Projektplan	29
3.1	Mitwirkung	30
4	Vergütung und Vertragsbedingungen.....	33
4.1	Erläuterung des Preisblattes / Angebotspreis	33
4.2	Erfolgsabhängige Vergütung und Vertragsgestaltung	33
4.3	Preisindikation Folgekosten	34
4.3.1	Betrieb	35
4.3.2	Lizenzen	37
4.3.3	Support	37
4.3.4	Externe Dienste	38
4.4	Reisekosten.....	40
4.5	Bindung an das Angebot	40
4.6	Zahlungsweise	40
4.7	Escrow-Agreement	41
5	Die Smart Media Engine	42
6	Unternehmensdarstellung	45
6.1	Überblick.....	45
6.1.1	Geschäftsbereiche	45
6.1.2	Fokus: KI-basierte Analyse und Kategorisierung von Content.....	46
6.1.3	Qualitätsmanagement.....	47
6.1.4	Datenschutz und Datensicherheit	47
6.2	Projektrelevante Referenzen.....	48
6.2.1	KI-basierte Wissensplattform im Umweltbereich (BMVI)	48
6.2.2	Semantische Analyse und Anreicherung von Datenbeständen (GVL)	49
6.2.3	Weitere relevante projektrelevante Referenzen im Überblick	50

7	Teil B: Erläuterung der Kriterien aus der Leistungsbeschreibung.....	51
7.1	Funktionale Anforderungen an das KI-Tool.....	52
7.2	Allgemeine und nicht-funktionale Anforderungen.....	59

1 Ausgangssituation und Projektbeschreibung

1.1 Der Auftraggeber

Die Landesanstalt für Medien NRW (medienanstalt-nrw.de) mit Sitz in Düsseldorf ist eine Anstalt des öffentlichen Rechts und die zuständige Aufsichtsbehörde für den privaten Rundfunk und Telemedien in Nordrhein-Westfalen.

Neben der Aufsicht über private Rundfunkanbieter und Telemedien hat die Landesmedienanstalt NRW (kurz auch: LfM) unter anderem die Aufgabe, die Medienkompetenz zu fördern und die Verbreitung, Nutzung und Fortentwicklungen von Medienangeboten zu beraten und durch wissenschaftliche Untersuchungen zu unterstützen.

Im Rahmen ihrer Kontrollfunktion ist die Landesanstalt für Medien NRW außerdem gesetzlich mit der Überwachung von Internetangeboten betraut und damit auch dafür zuständig, die Einhaltung des Jugendmedienschutzes zu prüfen.

Gesetzliche Grundlagen der Arbeit der Landesanstalt für Medien NRW sind das Landesmediengesetz Nordrhein-Westfalen (LMG NRW) und u. a. der Rundfunkstaatsvertrag (RStV). In Bezug auf jugendmedienschutzrechtliche Fragestellungen ist der Jugendmedienschutz-Staatsvertrag (JMStV) maßgeblich.

1.2 Gegenstand des Angebotes

Die Landesanstalt für Medien NRW möchte zur Unterstützung ihrer gesetzlichen Aufsichtstätigkeit ein IT-Tool zur Recherche und Identifikation von potenziell rechtswidrigen Inhalten entwickeln lassen. Hierfür hat die LfM Anfang 2020 ein Projekt für die Konzeption und Entwicklung eines solche Tools ausgeschrieben.

Das zu entwickelnde Tool soll die Recherche von potenziell rechtswidrigen Inhalten durch das „Monitoring-Team“ der Abteilung Recht & Technik der LfM vereinfachen und effizienter gestalten.¹

Derzeit ist die Recherche zeitintensiv und wird von den Mitarbeiterinnen und Mitarbeitern des Monitoring-Teams zum großen Teil händisch vorgenommen. Das Tool soll durch den Einsatz von Algorithmen, KI-Verfahren und Suchtechnologien potenziell rechtswidrige Inhalte im Internet finden. Dazu sollen (teil-) automatisiert Texte, Bilder, Audiofiles und Videos recherchiert und in ihnen potenziell rechtswidrige Inhalte identifiziert werden. Die verwendeten Algorithmen sollen im Laufe der Zeit durch nachträglich eingepflegte Trainingsdaten und Feedback des Monitoring-Teams immer treffgenauere Ergebnisse liefern.

Maßgeblich für die von der LfM zu identifizierenden und zu behandelnden Rechtsverstöße ist der Jugendmedienschutz-Staatsvertrag, der grob in unzulässige Angebote und entwicklungsbeeinträchtigende Angebote unterscheidet (§ 4, 5 JMStV). Hierunter fallen Angebote mit Inhalten wie Volksverhetzung, Kriegsverherrlichung, Verwendung von Nazisymbolen bzw. allg. Kennzeichen verfassungswidriger Organisationen, Leugnung des Holocaust, Gewalt- und Drogenverherrlichung, Selbstgefährdung und (in unterschiedlichen Schweregraden) erotische und pornografische Inhalte.

Gegenstand des vorliegenden Angebotes der Condat AG ist die Entwicklung des ausgeschriebenen IT-Tools.

1.3 Die Änderungen im überarbeiteten Angebot

Das hier vorliegende Angebot ist unsere finale Überarbeitung des Angebots vom 07.04.2020.

Zur besseren Übersicht haben wir Absätze, die wir geändert oder neu geschrieben haben, mit einer grauen Markierung unterlegt. Änderungen bzw. gänzlich neue Abschnitte hat es an den folgenden Stellen gegeben:

In Ausgangssituation und Projektbeschreibung

Neue hinzugefügte Abschnitte:

- 1.3 Die Änderungen im überarbeiteten Angebot

In Lösung und Leistung

Neue hinzugefügte Abschnitte:

- 2.2.5 Black- und Whitelisting von Anbietern und Quellen
- 2.3.5 Integration externer Dienste
- 2.3.6 Abschließende Bewertung und Kategorisierung

Geänderte Abschnitte (punktuelle Anpassungen und Ergänzungen):

- 2.2.1 Automatisierte, periodische Internet-Suchen
- 2.2.4 Multimediale Quellen
- 2.3.2 Semantische Analyse
- 2.3.3 Klassifikation und Identifikation
- 2.3.4 Kontinuierliche Verbesserung - das lernende System
- 2.4.1 Datenbank der identifizierten Inhalte
- 2.4.2 Workflow zur Abarbeitung der identifizierten Inhalte
- 2.5.2 Analyse & Identifikation

In Vorgehensvorschlag und Projektplan

Neue hinzugefügte Abschnitte:

- 3.1 Mitwirkung

Änderungen:

- Ergänzung Evaluation externer Dienste
- Aussagen zu möglichem späteren Projektstart
- Ergänzung des Projektablaufplans (Gantt-Chart) um benötigten Grad der Mitwirkung seitens der LfM NRW

In Vergütung und Vertragsbedingungen

Neue hinzugefügte Abschnitte:

- 4.2 Erfolgsabhängige Vergütung und Vertragsgestaltung
- 4.7 Escrow-Agreement

Änderungen:

- 4.1 Erläuterung des Preisblattes / Angebotspreis
- 4.3 Preisindikation Folgekosten: Vertragskonditionen für Betrieb, Lizenzen und Support
- 4.3.2 Lizenzen: Anpassung des Lizenzpreises für die Plattform im Zuge der Risikoverlagerung auf die Condat
- 4.5 Bindung an das Angebot: Anpassung Bindefrist
- 4.6 Zahlungsweise

Mitgeltende Dokumente

- In der Anlage zu diesem Dokument übermitteln wir eine angepasste Fassung des Preisblatts. Vgl. auch die Erläuterungen in Abschnitt 4.1.
- Alle anderen Anlagen und mitgeltenden Dokumente unseres ursprünglichen Angebots 20-0702 sind unverändert und bleiben somit unverändert gültig.

2 Lösung und Leistung

2.1 Der Lösungsansatz im Überblick

Die Realisierung des IT-Tools erfolgt auf Basis der
der Condat.

Sie ermöglicht die bedarfsorientierte Integration geeigneter verfügbarer Services zur Analyse und Bewertung von Content,

Die Plattform gliedert sich in die Komponenten

- Recherche
- Analyse & Identifikation
- Bearbeitung und Verwaltung der Inhalte (Datenbank)

Recherche

Die Recherche-Komponente ermittelt durch die Anbindung allgemeiner Suchmaschinen
, dedizierter Such-APIs) und durch Crawlen von Internet-Angeboten Kandidaten für rechtswidrige Inhalte und übergibt diese zur weiteren Verarbeitung an die Komponente *Analyse & Identifikation*.

Analyse & Identifikation

Die Analyse & Identifikations-Komponente führt die Analyse und Bewertung der durch die Recherche-Komponente ermittelten potenziellen Inhalte mittels
und ggf. durch Nutzung geeigneter externer Services (s.o.) durch.

Inhalte, die als potenziell rechtswidrig eingestuft wurden, werden zur Speicherung und Bearbeitung an die Komponente *Datenbank* übergeben.

Bearbeitung und Verwaltung der Inhalte

Alle relevanten Informationen zu den identifizierten Inhalten werden strukturiert in einer Datenbank abgelegt und sind über ein such-basiertes UI zugreifbar.

In einer Task-Management-Komponente wird für jeden abzuarbeitenden Inhalte eine korrespondierende Aufgabe angelegt. Für die Abarbeitung der Aufgaben durch die Nutzer (i.e. das Monitoring-Team) ist ein reduziertes Workflow-UI bestehend Dashboard, Aufgabenlisten sowie Detail- und Bearbeitungsansicht in das UI der Content-Discovery-Plattform integriert. Das UI integriert die aufgabenbezogenen und die inhaltsbezogenen Informationen und Funktionen.

2.2 Recherche

2.2.1 Automatisierte, periodische Internet-Suchen

Hierbei ist entscheidend, dass in enger Zusammenarbeit mit dem AG die verwendeten Suchbegriffe definiert und auch ständig aktualisiert werden; dies erfolgt auf Basis der sog. Ontologie (s. dazu unten *Wissensmodell*).

In dem hier vorgestellten Ansatz soll dazu ein KI-Verfahren zum Einsatz kommen, das sich *Semantic Fingerprinting* nennt (mehr dazu s.u.). Diese ermöglichen ein *Content Filtering* Verfahren, das ähnlich einem Spam-Filter funktioniert, jedoch unter Anwendung semantischer Methoden; die dabei herausgefilterten URLs werden dann zur Beurteilung in die Watchlist gespeichert.

Optional können hier weitere Dienste angebunden werden,

In Bezug auf die Internet-Suche ist zu berücksichtigen, dass dabei nur Quellen gefunden werden können, die direkt zugreifbar und nicht zugriffsgeschützt sind (wie z.B. i.A. Foren).

Des Weiteren hängt die Aktualität der Suche immer von der Aktualität des Suchmaschinen-Indexes ab, sodass man i.d.R. nicht davon ausgehen kann, dass die Suchergebnisse immer tagesaktuell sind. Teilweise betragen die Indizierungsintervalle bis zu zwei Wochen.

Für die Ermittlung der Reichweite eines Angebots wird
bunden.

eingel

2.2.2 Spezialisierte Suchen in sozialen Netzwerken

Bei den sozialen Netzwerken lässt sich besonders einfach die Reichweite bestimmen, da Angaben wie Anzahl Follower/Views/Retweets über die APIs zumeist abrufbar sind. Auch hier wird anschließend eine Relevanzbewertung durchgeführt. ↯

2.2.3 Crawling individueller Quellen

Individuelle, bekannte Quellen (URLs) können direkt beobachtet, d.h. über einen Crawler analysiert werden.⁴

2.2.4 Multimediale Quellen

Die oben aufgeführten Ansätze funktionieren prinzipiell in gleicher Weise für textuelle wie für multimediale Inhalte (Videos, Audio/Podcasts, Bilder). Die Suchmaschinenanbieter stellen über ihre APIs spezielle Suchen nach Videos und Bildern zur Verfügung, bei den spezifischen sozialen Netzwerkdiensten (optionale Erweiterung) ist das ohnehin gegeben.

Zur Analyse von Audio- und Videoinhalten siehe Abschnitt 2.3.5 Integration externer Dienste.

Bei Bildern ist es einfach möglich, über eine sog. *Reverse Image Search* ähnliche Bilder direkt über den Bildinhalt im Internet zu finden, insbes. auch Bilder, die von anderen Sites wiederverwendet werden. Dies ist sinnvoll, wenn das Ursprungsbild schon als rechtswidrig erkannt wurde. Aus dem Forschungsprojekt ⁵ verfügt der Anbieter auf diesem Gebiet über umfangreiche Erfahrungen.⁶

2.2.5 Black- und Whitelisting von Anbietern und Quellen

Zur Verbesserung von Recall und Precision (Vermeidung von falsch-positiven Ergebnissen) wird ein Whitelisting und auch Blacklisting (bekannte Anbieter rechtswidriger Inhalte) unterstützt. Die Listen können einerseits direkt manuell verwaltet werden. Aber auch über einen Feedbackmechanismus aus dem Bearbeitungsworkflow aktualisiert werden.

2.3 Analyse und Identifikation der Inhalte

2.3.1 Wissensmodell

Um in großen Mengen von Dokumenten (Texte, Bilder, Audios, Videos) rechtswidrige Inhalte identifizieren zu können, müssen solche Inhalte geeignet definiert und die Dokumente inhaltlich analysiert werden. Hierfür verwenden wir externe und händisch erstellte Wissensquellen.

Das dabei entstehende Wissensmodell kann automatisch aktualisiert werden und der Begriffs-Bereich der rechtswidrigen Inhalte ist zur Erreichung einer größeren Genauigkeit händisch ergänzbar. Insbesondere werden bereits vorhandene Begriffskataloge der LfM NRW in die Wissensbasis integriert. Das Wissensmodell ist mehrsprachig und gestattet dadurch Analysen nicht nur für deutschsprachige Dokumente.

2.3.2 Semantische Analyse

Die textuellen Bestandteile der Dokumente, also Texte, Metadaten und ggf. in Text umgewandelte Audios und Video-Tonspuren, werden einer automatischen, umfangreichen sprachlichen und semantischen Analyse unterzogen.

Das Ergebnis ist für jedes Dokument, unabhängig von seinem Typ, ein sogenannter semantischer Fingerabdruck, d. h. eine gewichtete Liste von eindeutigen Konzepten und Entitäten, die seinen thematischen Inhalt beschreibt. Dabei werden insbesondere auch die Begrifflichkeiten, die auf rechtswidrige Inhalte hindeuten, erkannt.

Da die semantischen Fingerabdrücke dokumententyp-übergreifend sind, können mit ihnen unterschiedliche Arten von Dokumenten miteinander verglichen werden

Durch die Verwendung von sprachunabhängigen Konzepten mit Hilfe von sind die semantischen Fingerabdrücke außerdem sprachunabhängig und ermöglichen dadurch sprachübergreifende Suchen.

Das impliziert also auch eine automatische, sprachliche Erweiterung der Begriffe zu rechtswidrigen Inhalten.

Durch das Verfahren der Fragmentierung in de , können komplette Dokumente, z. B. Webseiten, oder auch Teile davon einzeln analysiert und bewertet werden.

Für das IT-Tool der LfM NRW wird die semantische Analyse der zur Erfüllung der spezifischen fachlichen Anforderungen im Wesentlichen in zwei Bereichen weiterentwickelt :

- Erkennung von Phrasen ("Mit deutschem Gruß")
- Language Processing von Social Media Inhalten

Die Analyse von Inhalten in Bildern und Videos erfolgt über die Integration von externen Diensten. Siehe dazu Abschnitt 2.3.5 Integration externer Dienste.

2.3.3 Klassifikation und Identifikation

Für die Identifikation der unterschiedlichen Arten von rechtswidrigen Inhalten werden in der **_____** sogenannte Profile definiert. Jedes Profil beschreibt eine Art rechtswidrigen Inhalt („Hasssprache“, „Pornographie“, usw.). Profile besitzen ebenfalls einen **_____** semantischen Fingerabdruck, der die Art des rechtswidrigen Inhaltes beschreibt. Üblicherweise werden für ihre Definition mehrere Beispiel-Dokumente mit dem jeweiligen rechtswidrigen Inhalt verwendet.

Damit können Konzepte und Entitäten entfernt oder hinzugefügt werden bzw. deren Relevanz verändert werden. Wichtig hierbei ist, dass beschreibende Stichwörter und Begrifflichkeiten automatisch auch auf Konzepte abgebildet werden. Lediglich bei mehrdeutigen Begriffen ist vom Nutzer eine Auswahl zu treffen. Weiterhin bekommt das Profil noch den Typ des rechtswidrigen Inhalts als Namen. Da die Fingerabdrücke der Dokumente sprachenunabhängig sind, sind dies die Fingerabdrücke der Profile automatisch auch, d. h. die rechtswidrigen Inhalte sind für alle relevanten Sprachen definiert. Schließlich können auch neue Profile angelegt und (nicht mehr verwendete) Profile gelöscht werden.

Ausgehend von einem Profil können Dokumente mit einem entsprechenden rechtswidrigen Inhalt gesucht werden oder für ein analysiertes Dokument kann ermittelt werden welche rechtswidrigen Inhalte vorliegen. Dies können durchaus auch mehrere sein.

Für das IT-Tool der LfM NRW wird die Klassifikation und Ähnlichkeitssuche der **_____** zur Erfüllung der spezifischen fachlichen Anforderungen um die Möglichkeit erweitert, Konzepten im semantischen Fingerabdrücken ein negatives Relevanzmaß zuzuordnen. Damit wird die Vermeidung von falsch-positiven Ergebnissen verbessert (z.B. rechtsextremer Inhalt vs. Bericht über Rechtsextremismus).

Die Suchprozesse sind automatisch im Hintergrund, auch wiederholt, ausführbar oder vom Nutzer einzeln anstoßbar. Die identifizierten Inhalte werden mit einer zusätzlichen Prozess-Id versehen und in einer internen Datenbank gespeichert. Diese kann über einfache Schnittstellen ausgelesen werden. Zusätzlich wird noch der Zeitpunkt der Identifikation gespeichert. Dies ist sinnvoll, da sich die Dokumente in ihrer externen Quelle jederzeit ändern können. Für Bilder und Videos ist die Speicherung eines Thumbnails vorgesehen.

Ein großer Vorteil des Verfahrens besteht darin, dass die Ergebnisse erklärt werden können. Alle Schritte können in allen Einzelschritten komplett nachvollzogen werden und die entsprechenden umfangreichen Konfigurationsparameter angepasst werden.

2.3.4 Kontinuierliche Verbesserung - das lernende System

Alle Analyse-Algorithmen für Text, Ton und ggf. Grafik sind bereits mit initialen Daten und Trainings-Datensätzen ausgestattet, die von Condat für die speziellen Anforderungen der LfM NRW ergänzt werden. Hierzu werden dann auch die Expertise und Zuarbeit des Monitoring-Teams benötigt. Insbesondere werden die bereits bestehenden Stichwortlisten zur Definition der rechtswidrigen Inhalte als Grundlage zur Definition der Profile verwendet.

Die semantischen Fingerabdrücke der Profile werden durch die Bestätigung von korrekt erkannten rechtswidrigen Inhalten kontinuierlich verbessert. Hierfür stellt die Schnittstellen zur Verfügung, die lediglich mit den notwendigen Parametern aufgerufen werden. Mit Hilfe der oben bereits beschriebenen Merge-Prozesse werden dann die Semantischen Fingerabdrücke der Profile erweitert, um rechtswidrige Inhalte zukünftig noch besser erkennen zu können. Zusätzlich werden die Verstöße und ihre wichtigen Kenndaten für statistische Auswertungen gespeichert.

Auch die grundlegende Wissensbasis wird automatisch aktualisiert, wobei dies allerdings direkt keine Veränderung der Profile und Analyseergebnisse der bereits bekannten Ergebnisse bewirkt. Es ist aber administrativ möglich die Analyse der Dokumente und das Mergen der Profil-Fingerabdrücke komplett zu wiederholen, sollte sich die Wissensbasis signifikant verändern. Dies geschieht sinnvollerweise über Nacht.

Neben der kommen auch ML-basierte Verfahren (s.u. Abschnitte 2.3.5 und 2.3.6) zur Anwendung, die ebenfalls durch die Bestätigung von korrekt erkannten rechtswidrigen Inhalten kontinuierlich verbessert werden.

2.3.5 Integration externer Dienste

Zur Erreichung der erforderlichen Präzision bei der Unterscheidung zwischen potenziellen Verstößen und aus Aufsichtssicht unproblematischen Inhalten erscheint die Integration von Machine-Learning-basierten Verfahren insbesondere für die Analyse von multimediale Inhalten unabdingbar.

Dazu werden primär verfügbare externe Dienste integriert („Low-Code-Ansatz“). Das individuelle Entwickeln und Trainieren von neuronalen Netzen o.ä. ist einerseits aus Kostengründen im Rahmen dieses Projektes nicht möglich. Andererseits bieten externe Dienste auch den Vorteil, von ihrer zu erwartenden Weiterentwicklung durch den Anbieter profitieren zu können.

Analyse von Video-Inhalten

Für die Verarbeitung von Video-Inhalten wird ein externer Dienst integriert

Analyse von Bildern

Für die Analyse von Bildern bieten die großen Cloud-Anbieter entsprechende Dienste an, die teilweise u.a. auch „anstößige“ Inhalte in

Bildern erkennen sollen. Weitere in Frage kommende Dienste/Anbieter sind

Zu Beginn des Projekts wird eine Evaluation dieser Dienste durchgeführt und ermittelt, welcher Dienst am besten für das IT-Tool geeignet ist.

Unabhängig davon kann bereits heute davon ausgegangen werden, dass diese Dienste insbesondere bei der Analyse von Bildern, die Nacktheit („Nudity“) enthalten, keine hinreichende Unterscheidung zwischen aus Aufsichtssicht nicht zu beanstandenden Bildern und rechtswidrigen Inhalten liefern werden. Für die beiden von der LfM NRW bereitgestellten Beispiele von nicht zu beanstandenen Bildern liefern diese Dienste bereits die „höchste“ Kategorie „raw nudity“. Dieses Ergebnis kann höchstens dazu dienen, in einem stufenweisen Verfahren sicher „harmlose“ Bilder von solchen zu unterscheiden, die weiter analysiert werden müssen.

Zur Lösung dieser Problematik ist die Einbindung des vorgesehen, der das selbstständige Trainieren des Cloud-Dienstes mit eigenen Bildern erlaubt.

Kostenkontrolle

Die o.g. externen Dienste sind in der Regel kostenpflichtig. Eine Abschätzung dieser Kosten setzt eine Kenntnis des zu erwartenden Mengengerüsts voraus, die aktuell noch nicht gegeben ist.

Um prinzipiell im Betrieb, falls notwendig, eine Beschränkung bzw. Kontrolle dieser Kosten zu ermöglichen, werden entsprechende Mechanismen und Konfigurationsmöglichkeiten in das IT-Tool integriert:

2.3.6 Abschließende Bewertung und Kategorisierung

Die Erweiterung der Analyse um externe Dienste zur Verarbeitung von multimedialen Inhalten ergibt, dass zu einem Content, z.B. einer HTML-Seite mit mehreren Bildern oder einem Video „zerlegt“ in Keyframes und Transskript, mehrere Analyseergebnisse (textuelle und semantische Analyse, Ergebnisse externer Dienste) vorliegen, die zusammen mit weiteren Metainformationen (z.B. Anbieter, Quelle) in ihrer Gesamtheit zu bewerten sind.

Dazu wird ein einfaches Verfahren implementiert, das mittels überwachtem Lernen initial zu trainieren ist und während des Betriebs des IT-Tools durch die Verarbeitung von User-Feedback kontinuierlich lernt.

2.4 Bearbeitung und Verwaltung der Inhalte (Datenbank)

Inhalte, die durch die Komponente *Analyse & Identifikation* als potenziell rechtswidrig eingestuft wurden, werden zur Speicherung und Bearbeitung an die Komponente *Datenbank* übergeben.

2.4.1 Datenbank der identifizierten Inhalte

In der *Datenbank* werden alle relevanten Informationen zu den identifizierten Inhalten inkl. des Bearbeitungsstatus strukturiert abgelegt. Hierfür wird die dokument-orientierte Datenbank verwendet.

Neben den Inhalten werden auch identifizierte Anbieter und Angebote strukturiert abgelegt und verwaltet, so dass eine entsprechende Zuordnung und Auswertung möglich ist.

Für den Zugriff auf die in der *Datenbank* gespeicherten Inhalte steht ein UI zur Verfügung, mit dem die Inhalte nach Merkmalen gefiltert werden können.

Mit dem Tool werden Statistiken über die Inhalte der Datenbank erzeugt und visualisiert.

2.4.2 Workflow zur Abarbeitung der identifizierten Inhalte

Die *Datenbank*-Komponente beinhaltet eine *Task-Management*-Komponente, in der für jeden abzuarbeitenden Inhalte eine korrespondierende Aufgabe angelegt wird. Zusammenhängende Inhalte können auch in einer Aufgabe geeignet gruppiert werden, z.B. durch Unteraufgaben je Inhalt.

Die Aufgaben erhalten alle relevanten Attribute des zugeordneten Inhalts, die für die Sortierung, Filterung, Priorisierung und Zuweisung der Aufgaben notwendig sind.

Die initiale Konfiguration der Regel und auch mögliche Werte für Prioritäten sind in der *Task-Management*-Komponente konfigurierbar. Die Konfiguration erfolgt im Rahmen des

Projekts in Abstimmung mit dem AG. Die Regel kann auch im späteren Betrieb durch den AG über das Admin-UI angepasst werden.

Die Aufgaben-Id in der Task-Management-Komponente fungiert als eindeutiger Schlüssel für den Verstoß. Zu einem Inhalt (ggf. Angebot als Menge von Inhalten) gibt es zu einem gegebenen Zeitpunkt immer nur eine aktive Aufgabe. Im Laufe der Zeit kann es (z.B. bei wiederholten Verstößen) zu einem Inhalt mehrere Aufgaben geben, die dann auch jeweils eigene Ids haben.

Für die Abarbeitung der Aufgaben durch die Nutzer (i.e. das Monitoring-Team) ist ein reduziertes Workflow-UI

in das UI der -Plattform integriert.

Die Detail- und Bearbeitungsansicht integriert die aufgabenbezogenen Informationen und Funktionen der Task-Management-Komponente (Kommentare, Anhänge,

Statusänderungen, Bearbeiter zuweisen) und die inhaltsbezogenen Informationen und Funktionen in einer Ansicht.

Dem Nutzer werden hier die Analyseergebnisse angezeigt und er kann bei Bedarf auch weitergehende Analyseschritte starten, z.B. die Sprach- und Bildanalyse für ein Video, das nur aufgrund seiner textuellen (Meta)daten als potenziell rechtswidrig eingestuft wurde.

Zur Dokumentation eines Rechtsverstoßes kann der Nutzer eigene Dateien (z.B. Screenshots) hochladen, als auch Dokumente aus in der Datenbank gespeicherten Analyseergebnissen automatisch generieren lassen,

2.4.3 Benutzerverwaltung

Für die Verwaltung der Benutzer der *Datenbank*-Komponente wird ebenfalls die Software verwendet.

2.5 Architektur



Die technischen Grundlagen der Komponenten Recherche und Analyse & Identifikation werden im Folgenden detaillierter beschrieben.

2.5.1 Recherche-Komponente

Die Recherche-Komponente enthält in der Konfiguration, die für das Recherchetooll zum Einsatz kommen würde, die im Folgenden aufgeführten Microservices (die teilweise noch konfiguriert und im Rahmen der Umsetzung aus bestehenden Komponenten abgeleitet werden müssen):



Bei Bedarf können hierbei auf einfache Weise weitere Agenten als Microservices eingebunden werden.

2.5.2 Analyse & Identifikation

Auf die gleiche Weise, wie die eben beschriebenen internen Dienste in den Workflow integriert sind, können auch spezielle externe Dienste eingebunden werden. 1

3 Vorgehensvorschlag und Projektplan

Für den Projektverlauf schlagen wir – nach aktuellem Kenntnisstand – das im Folgenden beschriebene Projektvorgehen vor.

Der grundsätzliche Ansatz ist, dass die Software in kurzen Iterationen (sog. „Sprints“) in enger Abstimmung mit dem Kunden agil entwickelt wird. Jeder Sprint dauert drei Wochen. Am Ende eines jeden Sprints stellen wir ein lauffähiges System, das genutzt wird, um mit dem Auftraggeber entlang des Projektfortschritts die Umsetzung der Anforderungen zu überprüfen und Anforderungen ggf. nachzuschärfen bzw. Korrekturen vorzunehmen.

Im Einzelnen schlagen wir vor:

- In einer ersten dreiwöchigen Konzeptionsphase werden die Anforderung des Auftraggebers noch einmal im Detail aufgenommen, offene Fragen geklärt und die technischen Systeme in der Entwicklungsumgebung bereitgestellt.
- Im Anschluss werden im ersten Sprint („Sprint 0“) die Erhebung und Abstimmung der Anforderungen abgeschlossen und das daraus entwickelte Konzept zur Umsetzung finalisiert. Am Ende von Sprint 0 setzen wir eine erste Version des Systems auf, das zu diesem Zeitpunkt im Kern noch dem Standard der entspricht.
- **Parallel werden in Sprint 0 mögliche externe Dienste in Hinblick auf ihre Eignung für das IT-Tool evaluiert.**
- Die eigentliche Software-Entwicklung bzw. das Customizing des Systems entlang der Kunden-Anforderungen erfolgt im Anschluss in vier dreiwöchigen Sprints. Am Ende eines jeden Sprints steht eine neue Version des IT-Tools bereit. Die Software-Entwicklung wird jeweils qualitätsgesichert. Anhand der jeweils vorliegenden Systemversion werden die Anforderungen in Abstimmung mit dem Auftraggeber aktualisiert bzw. fortgeschrieben.
- In einer vierwöchigen Phase wird dann das IT-Tool in Betrieb genommen und dem Kunden bereitgestellt. Die Dokumentation wird erstellt und die Inhaus-Schulung für die Mitarbeiter durchgeführt, die später mit dem IT-Tool arbeiten werden. Dabei wird auch das Wissen vermittelt, wie das System mit Inhalten gefüllt und trainiert werden kann.
- In der letzten Phase wird das IT-Tool vom Auftraggeber in den Pilotbetrieb übernommen. Bei diesem Go-Live unterstützen wir den Auftraggeber beim Integrieren der ersten Daten, bei der inhaltlichen Konfiguration und bei den ersten Durchläufen des Trainings des Systems.
- Parallel zu allen Aktivitäten wird das Projekt von einem erfahrenen Projektmanager gesteuert, der den Projektfortschritt managt und Entscheidungen, die für den Projektverlauf relevant sind, mit dem Auftraggeber herbeiführt.

3.1 Mitwirkung

Für den Projekterfolg ist eine aktive Mitwirkung des Auftraggebers im Rahmen des Projektverlaufs unabdingbar.

Der Umfang dieser Mitwirkung kann zum jetzigen Zeitpunkt nur grob geschätzt werden. Wir gehen dabei von folgenden Umfangskategorien aus:

- **Geringer Aufwand:** Punktuelle Mitwirkung einzelner Mitarbeiterinnen²³ des Auftraggebers, z.B. für Einzelauskünfte, Priorisierungen oder Entscheidungen. Der Aufwand bewegt sich hier jeweils im Bereich von Minuten bis zu wenigen Stunden.
- **Mittlerer Aufwand:** Partiiell stärkere Mitwirkung einzelner Mitarbeiterinnen, z.B. bei der Teilnahme an Sprint-Review Sitzungen inkl. der Vor- und Nachbereitung.

²³ Der sprachlichen Einfachheit halber verwenden wir im Folgenden durchgängig die weibliche Form für Mitarbeiter*innen jeden Geschlechts.

Der Aufwand bewegt sich für die beteiligten Mitarbeiterinnen im Bereich von mehreren Stunden bis zu zwei Tagen pro Woche.

- **Hoher Aufwand:** Umfangreiche Mitwirkung von mehreren Mitarbeiterinnen des Auftraggebers, z.B. beim Trainieren des IT-Tools oder bei Workshops für die Anforderungsermittlung. Der Aufwand bewegt sich im Bereich von mehreren Tagen pro Mitarbeiterin pro Woche und insg. bis zu 50% der Kapazität des beteiligten Teams.
-

4 Vergütung und Vertragsbedingungen

4.1 Erläuterung des Preisblattes / Angebotspreis

Die Preisangaben für die in diesem Angebot vorgesehenen Arbeiten zur Entwicklung des ausgeschriebenen *IT-Tools zur Recherche potenziell rechtswidriger Inhalte im Internet* finden sich im Preisblatt, das von uns zusammen mit diesem Angebots-Dokument über das Biertool Cosinex an den Auftraggeber übermittelt wurde.

Das Preisblatt entspricht im Prinzip dem Preisblatt, das der Auftraggeber im ursprünglichen Vergabeverfahren zur Verfügung gestellt hatte. Für das finale und überarbeitete Angebot mussten wir die Tabelle im Preisblatt leicht anpassen und im unteren Ende erweitern, um die erfolgsabhängige Vergütung (s. folgenden Abschnitt) transparent ausweisen zu können.

4.2 Erfolgsabhängige Vergütung und Vertragsgestaltung

Sowohl im Bietergespräch als auch im Anschreiben zur Aufforderung für die finale Angebotsabgabe hat der Auftraggeber nachgefragt, „inwiefern erfolgsabhängige Faktoren in die Vertragsgestaltung mit aufgenommen werden können.“

Wir hatten dazu im Pitchgespräch bereits unsere prinzipielle Bereitschaft erklärt. Weil wir fest von unserer Leistungsfähigkeit und dem von uns vorgeschlagenen Ansatz überzeugt sind und uns sicher sind, dass wir für Sie eine Lösung entwickeln werden, die Sie bei der Bewältigung der Aufgaben rund um die Identifikation potenziell rechtswidriger Inhalte wirkungsvoll unterstützen wird.

Dazu bieten wir folgende Preis- und Vertragsgestaltung an

- In den ersten sechs Monaten nach dem Go-Live der Lösung räumt Condat dem Auftraggeber das Recht ein, den Vertrag zum Betrieb des Systems inkl. Lizenzen monatlich vorzeitig zu kündigen, sollte er mit der Qualität der Projektergebnisse grundsätzlich nicht zufrieden sein. Dies reduziert für den Auftraggeber das Risiko,

Dieser Vorschlag bringt zwar weiterhin keine vollständige Erfolgsgarantie. Aber: Aus unserer Sicht ist aufgrund der Kombination von vorzeitiger Kündigungsmöglichkeit während der sechsmonatigen Probezeit und den reduzierten initialen Projektkosten der Druck auf die Condat zur Schaffung einer qualitativ überzeugenden Lösung gestiegen und das Risiko des Auftraggebers gleichzeitig signifikant reduziert.

Ergänzend möchten wir nochmal bekräftigen, dass die Anforderungen an das IT-Tool sich in weiten Teilen mit der strategischen Roadmap für die [redacted] decken, so dass ein Teil der notwendigen Entwicklungen an der [redacted] nicht dem Projektbudget zugeschlagen wurde, sondern von der Condat im Zuge der Produktweiterentwicklung als strategisches Investment eingebracht wird.

Dieses gilt insbesondere auch für die von uns neu vorgesehene Evaluation externer Dienste²⁴

[redacted] für den Auftraggeber kostenneutral durchführen werden.

4.3 Preisindikation Folgekosten

Laut Angaben des Auftraggebers (vgl. Antworten auf die Bieterfragen) sollen die Kosten für den Betrieb, die Wartung/Support sowie für Lizenzen nicht in den Angebotspreis einberechnet, sondern separat und indikativ ausgewiesen werden. Die Details zu diesen Folgekosten werden in den nächsten Unterabschnitten erläutert. **Neu aufgenommen haben**

²⁴ Siehe Abschnitt 2.3.5 **Integration externer Dienste**.

wir einen Abschnitt zu den Folgekosten durch die Einbindung externe Dienste (Abschnitt 4.3.4).

Für Betrieb, Lizenzen und Support gehen wir vom Abschluss eines Vertrags mit einer Mindestvertragslaufzeit von 12 Monaten aus, mit einem monatlichen Sonderkündigungsrecht im ersten Jahr innerhalb der ersten sechs Monate (vgl. Abschnitt 4.2). Der Vertrag verlängert sich automatisch um ein Jahr, wenn er nicht zwei Monate vor Ablauf gekündigt wird.

4.3.1 Betrieb

Für den Betrieb des zu entwickelnden Systems sind wir zunächst davon ausgegangen, dass System vom AN betrieben wird. Die nachfolgend genannten Zahlen beruhen auf Erfahrungen aus anderen Kundenprojekten der Condat und gehen von einer Nutzung der aus.

Denkbar sind auch die Nutzung alternativer Cloud-Anbieter oder auch gänzlich andere Betriebsmodelle, z.B. ein Hosting mit dediziert eigener Hardware oder On-Premise bei der LfM NRW. Eine Festlegung müsste in Abstimmung zwischen dem AG und Condat spätestens im Rahmen der Konzeptionsphase erfolgen.

Ausgehend von einem Betrieb in gehen wir von folgendem Bedarf aus:

- Für die Umsetzung sind zwei Rechner vorgesehen (jeweils 4 Kerne und 16 MB RAM).
- Die umfangreichen Analyse-Arbeiten für die gecrawlten Seiten werden auf eine Extra-Maschine ausgelagert (Analyse-Maschine).
- Alle übrigen Programme und Daten befinden sich auf der "Basis-Maschine".
- Die Rechner benötigen unterschiedlich große SSD-Festplatten, wie unten angegeben.

Die Transferkosten werden vernachlässigt,

2nd-Level-Support Standard-SLA 8x5

Servicezeiten:

- Werktags – außer samstags, Heiligabend, Silvester: 9:00 – 17:00 Uhr.

Reaktionszeiten:

- Betriebsverhindernder Mangel: Innerhalb der Servicezeiten qualifizierte Rückmeldung spätestens 2 Stunden nach Eingang der Fehlermeldung.
- Betriebsbehindernder Mangel: Innerhalb der Servicezeiten qualifizierte Rückmeldung spätestens 1 Werktag nach Eingang der Fehlermeldung.
- Leichter Mangel: Spätestens nach 5 Werktagen innerhalb der Servicezeiten.

Erweiterte SLAs für den 2nd-Level-Support und auch ein SLAs für einen 3rd-Level-Support (siehe 4.3.3) können bei Bedarf angeboten werden.

²⁵ Alle Preisangaben verstehen sich netto zuzüglich der gesetzlichen Mehrwertsteuer.

4.3.2 Lizenzen

4.3.3 Support

Neben Betrieb und 2nd-Level-Support besteht grundsätzlich die Möglichkeit weitergehende Support-Leistungen zu vereinbaren:

SLAs für einen 3rd-Level-Support

Die Behebung von Fehlern im System (Software-Fehler) an sich ist durch die gesetzliche Gewährleistung abgedeckt. Die Gewährleistung umfasst jedoch keine garantierten Reaktions- und Bearbeitungszeiten. Diese können bei Bedarf zur Absicherung des Betriebs vereinbart werden.

²⁶ Alle Preisangaben verstehen sich zuzüglich der gesetzlichen Mehrwertsteuer.

Anwender-Support

Bei Bedarf kann ein Support-Kontingent für die Anwender des Systems, insbesondere im Hinblick auf die Konfiguration und das fortlaufende Trainings des Systems, vereinbart werden.

Weiterentwicklung

Für die kontinuierliche Weiterentwicklung des Systems, z.B. die Integration neuer externer Services, kann ein entsprechendes Kontingent für Software-Entwicklungsleistungen vereinbart werden.

Ein kombiniertes, individuelles Angebot von SLAs für einen 3rd-Level-Support, Anwender-Support und Weiterentwicklungs-Kontingent ist möglich.

4.3.4 Externe Dienste

In Abschnitt 2.3.5 haben wir im Details ausgeführt, dass zur Erreichung der erforderlichen Präzision bei der Unterscheidung zwischen potenziellen Verstößen und aus Aufsichtssicht unproblematischen Inhalten die Integration von Machine-Learning-basierten Verfahren notwendig sein wird, insbesondere für die Analyse von multimedialen Inhalten. Und dass es aus mehrerlei Gründen ratsam ist, hier auf einen Low-Code-Ansatz zu setzen und primär am Markt verfügbare externe Dienste zu integrieren.

Für einen Teil der aktuell vorgesehenen externen Dienste ist die Nutzung in der Regel kostenpflichtig. Eine Abschätzung dieser Kosten setzt eine Kenntnis des zu erwartenden Mengengerüsts voraus, die aktuell noch nicht gegeben ist. Auch ist noch nicht entschieden, welche Dienste im Hinblick auf den Einsatz im IT-Tool für die LfM NRW geeignet sind und final verwendet werden sollen.²⁷

Im IT-Tool für die LfM NRW planen wir verschiedene Mechanismen zu Kostenkontrolle und -steuerung beim Einsatz der externen Dienste zu implementieren.²⁸

²⁷ Condat wird diese Eignungsprüfung im Rahmen des Projektes auf eigene Kosten und für den AG kostenneutral durchführen.

²⁸ Vgl. hierzu die Passage zur Kostenkontrolle in Abschnitt 2.3.5.

Dies vorausgeschickt, weisen wir im Folgenden exemplarisch die Kosten für einige der möglichen externen Dienste aus.

²⁹ Alle Preis Stand 11.06.2020

4.4 Reisekosten

Im Rahmen des Projektes sind folgende Reisen zum Auftraggeber geplant und im Angebotspreis enthalten:

- Projektstart / Kickoff
- bis zu vier Konzeptions-Workshops (1-Tages-Workshops)
- Abschlusspräsentation
- Schulung

4.5 Bindung an das Angebot

Condat bindet sich bis zum _____ an dieses Angebot. Eine Verlängerung der Bindung muss schriftlich vereinbart werden.

4.6 Zahlungsweise

4.7 Escrow-Agreement

Auf Wunsch des Auftraggebers ist die Condat gern bereit, die Software der erstellten Lösung und der darin eingesetzten _____ im Rahmen eines Escrow-Agreements zu hinterlegen. Die Kosten hierfür trägt der Auftraggeber.

Die : wurde ursprünglich für die Unterstützung in den Redaktionen von TV- und Rundfunkanbietern entwickelt. Die eingesetzten Konzepte und Technologien sind aber so allgemein und umfangreich, dass die mittlerweile in einer Vielzahl von Branchen und Fachdomänen zum Einsatz kommt.

Die wichtigsten Features der **Smart Media Engine** im Überblick:

KI-basierte Analyse

Die KI-basierte, semantische Analyse der **Smart Media Engine** erfasst die Bedeutung von Inhalten und hilft, sie genau zu definieren und zu klassifizieren. Hierbei werden eindeutig allgemeine Konzepte, Personen, Organisationen, Orte und Ereignisse erkannt.

Fingerprinting

Die entstehenden Semantic Fingerprints (semantischen Fingerabdrücke) sind medientyp-unabhängig. Dadurch sind alle Prozesse unabhängig von der Art des Inhalts möglich.

Mehrsprachigkeit

Alle Prozesse der **Smart Media Engine** arbeiten mehrsprachig, wodurch die Ergebnisse sprachunabhängig sind.

Weitere Sprachen sind in der Regel leicht ergänzbar.

Wissen der Welt

Durch die Verwendung von Internet-Wörterbüchern sowie allgemeinen und spezifischen Wissensquellen beschreibt und beschreibt die Smart Media Engine Inhalte so genau und aktuell wie möglich.

Semantische Suche

Inhalte können mit der **Smart Media Engine** in natürlicher Sprache intuitiv und schnell gefunden werden, ohne enthaltene Schlüsselwörter oder Phrasen zu kennen.

Empfehlungen

Von der **Smart Media Engine** angereicherte Inhalte können Nutzern auf der Grundlage von Arbeitskontext und individuellen Präferenzen empfohlen werden.

Integration

Die **Smart Media Engine** ist nahtlos in das marktführende Newsroom-System OpenMedia der Firma SCISYS integriert und kann flexibel an weitere Tools angebunden werden.

Klassifikation

Die **Smart Media Engine** erkennt automatisch Themen und ordnet die Inhalte diesen Themen zu. Die Relevanz von Themen wird automatisch überwacht und dadurch Trends erkannt.

Personalisierung

Die **Personalisierung** kann Präferenzen von Nutzern erkennen und Such- und Empfehlungsergebnisse genau auf die Wünsche des jeweiligen Nutzers anpassen.

Aktualisierung

Die zugrundeliegende Wissensbasis wird von der **aktualisierung** automatisch und kontinuierlich aktualisiert und um neue Begrifflichkeiten erweitert, so dass die Ergebnisse aller Prozesse ständig aktuell sind.

Data Cleansing

Die **Data Cleansing** verfügt über umfangreiche Mechanismen zur Normierung und Bereinigung von Daten, oft ein essenzieller Bestandteil, der die weiteren High-End-Prozesse überhaupt erst möglich macht.

Anreicherung

Inhalte werden bei Bedarf von der **Anreicherung** um zusätzliche Informationen erweitert,

Smart News Gathering

Die **Smart News Gathering** kann Inhalte auch in sozialen Medien durchsuchen und die „heißesten“ Themen identifizieren. Der Meldungs-Stream wird dabei in Echtzeit analysiert und die aktuellen „Trending Topics“ geliefert.

Die **Smart News Gathering** in Fachpublikationen:

FKT / April 2020: "KI im Broadcast Newsroom – was gibt es heute schon?"

<https://www.fkt-online.de/news/news-detail/27638-ki-im-broadcast-newsroom-was-gibt-es-heute-schon/>

FKT / April 2019: "Externes Wissen: Neue, AI-basierte Methoden zur Klassifikation und Anreicherung von Inhalten"

<https://www.fkt-online.de/archiv/artikel/2019/fkt-4-2019/25546-externes-wissen-neue-ai-basierte-methoden-zur-klassifikation-und-anreicherung-von-inhalten/>

6 Unternehmensdarstellung

6.1 Überblick

Die Condat AG wurde vor 1979 von Absolventen der Technischen Universität Berlin gegründet. Unsere Leitidee, innovative Technologien, bewährte Standardprodukte und eigene Anwendungen mit hohem Kundennutzen zu kombinieren, wurde seitdem erfolgreich weiterentwickelt und bildet die Grundlage für zahlreiche erfolgreich durchgeführte Kundenprojekte. 2018 haben wir die Condat AG mit dem Internetdienstleister HFC Inter.net GmbH zusammengeführt und verfügen nun neben dem Hauptsitz in Berlin über einen zweiten Entwicklungsstandort im niedersächsischen Cloppenburg.

Zu unserem Dienstleistungsportfolio gehören

- Konzeption, Architektur- und Prozessentwicklung, Design
- Technische Realisierung
- Beratung, Einführung, Schulung
- Projekt- und Qualitätsmanagement
- Betriebsunterstützung, Wartung und Support

Mit der Mehrzahl unserer Kunden haben wir langjährig bestehende Geschäftsbeziehungen etabliert, in denen sich vertrauensvolle Partnerschaften entwickeln konnten.

6.1.1 Geschäftsbereiche

Die drei wesentlichen Branchen, in denen Condat über herausragende Erfahrungen verfügt, sind die **Medienindustrie, öffentlichen Verwaltungen sowie das Gesundheitswesen**. Trotz dieser Fokussierung sind insbesondere unsere eigenen Produkte so entwickelt worden, dass sie branchenübergreifend einsetzbar sind.

Für **Medienunternehmen** erstellen wir Online-Portale und IT-Lösungen auf der Basis von marktführenden Content Management Systemen und eigenen Produkten. Unsere Anwendungen unterstützen integrierte IT-Workflows bei Fernsehsendern mit den Schwerpunkten On-Demand-Angebote, Metadatenmanagement und Programm- und Sendeplanung.

Zu unseren Kunden zählen unter anderem ZDF, Deutschlandfunk, WDR, RBB und ARD Digital (siehe auch Referenzen in Kapitel 5.2).

Für die **öffentliche Verwaltungen** entwickeln wir hoch performante Online-Portale mit besonderen Anforderungen an Verfügbarkeit und Skalierbarkeit sowie individuelle Anwendungen zur Unterstützung von Fachverfahren u.a. im Umweltbereich. Zu unseren Kunden zählen unter anderem das Bundesministerium für Verkehr und digitale Infrastruktur, der Deutsche Bundestag, das Umweltbundesamt sowie das Bundesamt für Strahlenschutz (siehe auch Referenzen in Kapitel 5.2).

Für Unternehmen der **Gesundheitsbranche** entwickeln wir maßgeschneiderte Konzepte und IT-Lösungen für die Digitalisierung von Kommunikations- und Kollaborationsprozessen. Mit der **Smart Health Solution** haben wir eine modulare Produktplattform für das Gesundheitswesen entwickelt, die Ärzte, Patienten, Mitarbeiter und andere Gruppen digital kommunizieren und zusammenarbeiten lässt. Die Smart Health Solution ist unter anderem erfolgreich beim Städtischen Klinikum Braunschweig im Einsatz, das mit rund 65.000 stationären Patienten im Jahr zu den größten Häusern in Norddeutschland gehört.

6.1.2 Fokus: KI-basierte Analyse und Kategorisierung von Content

In den vergangenen Jahren hat sich unser Unternehmen zunehmend der Entwicklung von Anwendungen gewidmet, die multimedialen Content analysieren, bereinigen, um beschreibende Metadaten anreichern und nutzerspezifisch empfehlen können. Hintergrund sind vor allem die Anforderungen unserer Medienkunden, in deren (Nachrichten-)Redaktionen immer größere Mengen an Content für die Erstellung von kanalübergreifenden Beiträgen gesichtet werden müssen. Gleichzeitig steigt auch auf der Zuschauerseite der Bedarf, mit Hilfe von Empfehlungsalgorithmen individuelle, persönliche Medienerlebnisse zu gestalten.

Das durch zahlreiche erfolgreich umgesetzte Projekte generierte Wissen lassen wir kontinuierlich in die Entwicklung unseres Kernproduktes, der **Condat**, einfließen (weitere Informationen zum Produkt in Kapitel 6). Durch eine Vermarktungs-Kooperation mit CGI/SCISYS, dem Hersteller des führenden Newsroom-Tools OpenMedia, gleichen wir die angebotenen Funktionen regelmäßig mit den Anforderungen im Markt ab.

Gleichzeitig beteiligen wir uns seit vielen Jahren regelmäßig an nationalen und internationalen Forschungsprojekten, und integrieren dort erarbeitete innovative Konzepte und Verfahrensweisen in unser Produkt. Darüber hinaus unterhalten wir aktive Partnerschaften zu führenden Forschungseinrichtungen wie dem Deutschen Forschungsinstitut für Künstliche Intelligenz (DFKI).

6.1.3 Qualitätsmanagement

Das Qualitätsmanagement der Condat AG ist nach DIN EN ISO 9001:2000 zertifiziert und garantiert eine professionelle Projektabwicklung sowie die Erstellung von Software nach den höchstmöglichen Qualitätsstandards.

Für alle parallel zum Projekt verlaufenden QS-Maßnahmen (z.B. Code-Reviews) stellt Condat dem Kunden einen vom Projektteam unabhängigen QS-Beauftragten als Ansprechpartner zur Verfügung.

6.1.4 Datenschutz und Datensicherheit

Die Condat AG bekennt sich ausnahmslos zur Einhaltung des Datenschutzes im Sinne der EU-Datenschutzgrundverordnung und des Bundesdatenschutzgesetzes. Unseren Auftraggebern versichern wir, dass bei der Verarbeitung personenbezogener Daten die geltenden gesetzlichen Vorschriften eingehalten werden. Eine konsequente Anwendung der Prinzipien Datenschutz per Technikgestaltung bzw. durch datenschutzfreundliche Voreinstellung (Privacy by Design, Privacy by Default) sind für uns insbesondere vor dem Hintergrund unserer Lösungen für das Gesundheitswesen eine Selbstverständlichkeit.

Die Mitarbeiter der Condat sind in ihren Arbeitsverträgen auf das Datengeheimnis verpflichtet. Darüber hinaus ist es gute Praxis, dass Mitarbeiter der Condat auftragsbezogen auf Wunsch des Auftraggebers entsprechende Erklärungen zusätzlich abgeben.

Beim Thema Datensicherheit richtet sich die Condat nach den Empfehlungen des Bundesamtes für Sicherheit in der Informationstechnik (BSI).

In der Condat ist ein Mitarbeiter als Datenschutzbeauftragter benannt. Aktueller Datenschutzbeauftragter der Condat ist

6.2 Projektrelevante Referenzen

Die Condat AG verfügt über zahlreiche Referenzen, die ihre Leistungsfähigkeit in den Kernfragestellungen des vom LfM ausgeschriebenene Recherche-Tools dokumentieren:

6.2.1 KI-basierte Wissensplattform im Umweltbereich (BMVI)

Kunde: Bundesministerium für Verkehr und digitale Infrastruktur

Jahr der Fertigstellung: 2020

Partner: Deutsches Forschungszentrum für künstliche Intelligenz (DFKI)

Projektbeschreibung:

Bei der Planung und Genehmigung von Verkehrsinfrastrukturprojekten sind umfangreiche umwelt-, natur- und artenschutzrechtliche Anforderungen zu beachten, deren Umfang und Komplexität in den vergangenen Jahren stetig angestiegen sind. Um die am Planungs- und Genehmigungsprozess beteiligten Parteien bei der zeit- und ressourcenaufwändigen Recherche nach projektrelevanten Daten zu unterstützen, hat das Bundesministerium für Verkehr und Infrastruktur (BMVI) 2019 die Entwicklung einer softwarebasierten Wissensplattform initiiert, die unter Nutzung selbstlernender Technologien (künstliche Intelligenz) die Möglichkeit bietet, zu jeder fachlichen Frage im Umweltbereich den aktuellen wissenschaftlichen Stand abzurufen. Im Rahmen der Ausschreibung einer Machbarkeitsstudie mit Entwicklung eines korrespondierenden Demonstrators zur Realisierung der Plattform erhielt die Condat AG zusammen mit unserem Partner, dem Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI), den Zuschlag zur Umsetzung.

Im Ergebnis bestätigte die durchgeführte Machbarkeitsstudie und der erstellte Demonstrator, dass eine auf KI-Technologien basierende, selbstlernende Wissensplattform für das skizzierte Anwendungsszenario des BMVI umsetzbar und in besonderem Maße geeignet ist, um die Effizienz von Planungs- und Genehmigungsprozessen in der Verkehrswegeplanung zu erhöhen. Die in der Plattform eingesetzten Methoden der Künstlichen Intelligenz reduzieren den Rechercheaufwand im Vergleich zu einfachen, Keyword-basierten Suchen in verteilten Archiven signifikant und liefern deutlich genauere und umfassendere Ergebnisse. Die moderne und intuitiv nutzbar gestaltete Benutzeroberfläche der Plattform stellt vorhandenes, kontextbezogenes Wissen übersichtlich dar und erlaubt ein direktes Sichten der zugrundeliegenden Dokumente mit Sprungmarken zu relevanten Abschnitten.

6.2.2 Semantische Analyse und Anreicherung von Datenbeständen (GVL)

Kunde: Gesellschaft zur Verwertung von Leistungsschutzrechten mbH (GVL)

Jahr der Fertigstellung: 2016-2019

Projektbeschreibung:

Mit der Gesellschaft zur Verwertung von Leistungsschutzrechten verbindet uns eine mehrjährige Projekthistorie, in deren Verlauf wir große Datenbestände mit semantischen Methoden auf Basis der : analysiert, bereinigt, Metadaten angereichert und abgeglichen haben. In folgenden Themenbereichen waren wir für die GVL tätig:

Automatische Validierung von Künstlermitwirkungen

Im Projekt wurde ein Service implementiert, mit dessen Hilfe die Mitwirkung von Künstlern an Audio- und Videoproduktionen validiert werden kann. Der Service basiert auf NLP- (Natural Language Processing) und NER- (Named Entity Recognition) Funktionalitäten wie Textanalyse, Normierung, Synonymerkennung und Fuzzy-Verfahren sowie spezifischer Matching-Verfahren, die Bestandteil der sind. Einbezogen wurden unterschiedliche externe Wissensquellen, z.B. die deutsche Synchronkartei, fernsehserien.de und filmliib.de.

Semantische Anreicherung von Metadaten für Tonträger

Im Projekt wurde ein Service für die GVL implementiert und nachfolgend betrieben, der die externen Wissensquellen Discogs und Musicbrainz nutzt, um Tonträgerdaten zu bereinigen (Data Cleansing), Dubletten zu erkennen und mit zusätzlichen Daten anzureichern. Der Service basiert auf NLP- und NER-Funktionalitäten wie Textanalyse, Normierung, Synonymerkennung und Fuzzy-Verfahren sowie spezifischer Matching-Verfahren, die Bestandteil der sind.

Semantisches Matching von Klassik-Werken

Im Projekt wurde auf der Basis der externen Wissensquellen imslp.org, klassika.info und opera-arias.com eine Klassikwerke-Datenbank aufgebaut, um die Matching-Ergebnisse für Produkte mit Genre aus der klassischen Musik zu verbessern. Zum Einsatz kamen erneut NLP- und NER-Funktionalitäten wie Textanalyse, Normierung, Synonymerkennung und Fuzzy-Verfahren sowie spezifische Matching-Verfahren der .

6.2.3 Weitere relevante projektrelevante Referenzen im Überblick

Titel und Thema	Kunde	Jahr
Content Discovery Picture Desk Web-basierte Applikation zur Unterstützung der Recherche, Bearbeitung und Zuordnung von Bildmaterial zu Programm-Assets	Red Bee Media	2016-2017
InVID – In Video Veritas Verification-App für Nachrichtenagenturen, Fernsehsender und Newsportale zur Überprüfung der Authentizität von Videos aus Social Media	EU (Horizon 2020)	2016-2018
Qurator – Curation Technologies Entwicklung von innovativen Kuratierungs-Services auf Basis von KI und Semantischen Web-Technologien für Redakteure und Journalisten im TV-/Medienbereich	BMBF	2018-2021

7 Teil B: Erläuterung der Kriterien aus der Leistungsbeschreibung

Teil des Angebots (im Sinne eines mitgeltenden Dokumentes) ist die ausgefüllte Tabelle „Leistungsübersicht - IT Tool zur Recherche potenziell rechtswidriger Inhalte“, die Bestandteil der Vergabeunterlagen war. Die ausgefüllte Tabelle haben wir dem Auftraggeber zusammen mit unserem ursprünglichen Angebot vom 07.4.2020 übermittelt. Die dort dargestellten Sachverhalte sowie die folgenden Erläuterungen sind unverändert gültig.

Die geforderten Selbsteinschätzungen hinsichtlich des Level of Compliance und des Levels of Effort haben wir in der i -Tabelle ausgefüllt.

Da der Platz für die Kommentierung („Bemerkungen/Remarks“) in der -Tabelle für unseren Input zu knapp bemessen war, haben für sämtliche Anforderungen in den Blättern „KI-Tool Funct. Requ.“ und „General & Non-Funct. Requ.“ in diesem Teil B / Kap. 7 vorgenommen. Sie finden unten stehend für jede Anforderung einen eigenen Abschnitt, der jeweils die ID der Anforderung und den Anforderungstext in der Überschrift trägt.

Insgesamt möchten wir noch die folgenden Anmerkungen vorausschicken:

- In der Spalte Level of Effort der Leistungsbeschreibungs- i haben wir als „out of the box“ die Leistungen markiert, die wir der LfM NRW als Teil des Produktes ' ausliefern und die über den Lizenzpreis für abgedeckt sind, d.h. für die keinen projektspezifischen Entwicklungskosten anfallen.
- Für Anforderungen, zu deren Erfüllung eine „Entwicklung nötig“ oder „Konfiguration nötig“ ist, wurden die Aufwände abgeschätzt; die Aufwände sind im Angebotspreis enthalten (vgl. Kapitel 4). Eine Aufschlüsselung reichen wir bei Bedarf gerne nach – bzw. besprechen diese mit Ihnen im Rahmen der Angebotspräsentation.

7.1 Funktionale Anforderungen an das KI-Tool

01-001: Der Algorithmus muss potentiell rechtswidrige Inhalte in Texten, Bildern, Audio- und Videodateien erkennen können. Die höchste Priorität liegt zunächst auf der Analyse von Texten

Die Erkennung rechtswidriger Inhalte basiert auf der Erzeugung von sog. Semantischen Fingerabdrücken für einen analysierten Inhalt und einen semantischen Ähnlichkeitsvergleich mit den im System hinterlegten Profilen (semantischen Fingerabdrücken) für die verschiedenen Kategorien von Rechtsverstößen.

Die Semantischen Fingerabdrücke basieren auf textuellen Informationen. Für Bilder, Audio- und Videodateien werden initial die verfügbaren textuellen Informationen (Metadaten, ggf. Transkript und/oder Untertitel) herangezogen. Bei hinreichendem „Anfangsverdacht“ können, abhängig von konfigurierbaren Schwellwerten, weitergehende Analysen des Inhalts über externe Services (z.B. Speech-to-Text) angestoßen werden.

→ Teil A, Abschnitt 2.2.5 Black- und Whitelisting von Anbietern und Quellen

Zur Verbesserung von Recall und Precision (Vermeidung von falsch-positiven Ergebnissen) wird ein Whitelisting und auch Blacklisting (bekannte Anbieter rechtswidriger Inhalte) unterstützt. Die Listen können einerseits direkt manuell verwaltet werden. Aber auch über einen Feedbackmechanismus aus dem Bearbeitungsworkflow aktualisiert werden.

Analyse und Identifikation der Inhalte und Abschnitt 2.3.5 Integration externer Dienste

01-002: Alle gelabelten Datensätze, die zur Identifizierung eines Verstoßes geführt haben, müssen zu Trainingszwecken in das Tool zurückgeführt werden

Potentiell rechtswidrige Inhalte werden über Aufgaben in dem Workflow-UI den Nutzern zur Bearbeitung vorgelegt und können dort über eine entsprechende Workflow-Aktion bestätigt werden. Über einen internen, automatisierten Prozess werden alle so bestätigten Inhalte verarbeitet. Der Fingerabdruck des Inhalts wird verwendet, um die Definition des

Profils zu der erkannten rechtswidrigen Inhaltsart zu ergänzen, so dass vergleichbare Fälle in Zukunft noch besser erkannt werden.

→ Teil A, Abschnitte 2.3.4 **Kontinuierliche Verbesserung - das lernende System** und 2.4.2 **Workflow zur Abarbeitung der identifizierten Inhalte**

01-003: Der Algorithmus muss so aufgebaut sein, dass er die gefundenen potenziell rechtswidrigen Inhalte kategorisiert

Unterschiedliche Arten rechtswidriger Inhalte werden in dem System mit Hilfe von Profilen definiert. Der Erkennungsalgorithmus liefert automatisch auch die Art (Kategorie) des rechtswidrigen Inhalts.

→ Teil A, Abschnitt 2.3.3 **Klassifikation und Identifikation**

01-004: Kategorien, Keywords und Trainingsdaten müssen durch den Anwender flexibel erweiterbar sein.

In dem Admin-UI kann die Definition der rechtswidrigen Inhalte (semantische Fingerabdrücke von Profilen) bearbeitet werden. Diese können einzeln um Konzepte und Entitäten ergänzt werden oder es können solche entfernt werden. Auch die Gewichtung von Konzepten und Entitäten kann verändert werden. Weiterhin können komplette Inhalts-Dokumente einem Profil hinzugefügt oder aus ihm entfernt werden. Das System berechnet dann den Fingerabdruck des Profils entsprechend neu. Letzteres Verfahren ist wesentlich schneller, aber etwas ungenauer und eignet sich deshalb vor allem zu Beginn des Projekts für die initiale Konfiguration.

Für die Klassifikation von Bildern und die abschließende Bewertung der Analyseergebnisse kommen teilweise auch ML-basierte Verfahren zum Einsatz, die über dedizierte UIs trainiert werden können.

→ Teil A, Abschnitt 2.3.3 **Klassifikation und Identifikation**

→ Teil A, Abschnitt 2.3.5 **Integration externer Dienste**

→ Teil A, Abschnitt 2.3.6 **Abschließende Bewertung und Kategorisierung**

01-005: Der Algorithmus muss Usern die Möglichkeit bieten, die Ausführung eines Algorithmus jederzeit zu unterbrechen und wieder zu starten zu können

Über das Admin-UI der Recherche-Komponente können die Status aller Such- und Crawling-Prozesse eingesehen und laufende Prozesse ggf. pausiert werden.

→ Teil A, Abschnitt 2.2 Recherche

→ Kriterium 02-007: *Das Zielsystem muss mit mehreren Anwendern parallel nutzbar sein (unter Berücksichtigung der Performance)*

01-006: Das Tool muss die Möglichkeit bieten, dass Nutzer manuell URL's (klassische Websites, Blogs, Foren und Plattformen) für eine Durchsuchung angeben können

Über das Workflow-UI können durch die Nutzer auch manuell Aufgaben zur Bewertung von Inhalten erstellt werden. Beim Erstellen der Aufgabe spezifiziert der Nutzer die URL des zu analysierenden Inhalts. Das Anlegen der Aufgabe erzeugt automatisch den Crawling- und Analyseprozess (ggf. zeitgesteuert) und aktualisiert die angelegte Aufgabe automatisch mit den Ergebnissen.

URLs für das Crawling können auch im Admin-UI der Recherche-Komponente manuell hinzugefügt werden.

→ Teil A, Abschnitt 2.2 Recherche

01-007: Das Tool muss die Möglichkeit bieten, automatisch in definierten Zeiträumen nach den gewünschten Inhalten zu suchen / zu crawlen. Die zu durchsuchenden URLs müssen automatisch generiert werden

Für alle Recherche-Prozesse (automatisierte Suchen, Crawling) können die Zeiträume, auch periodische, über das Admin-UI der Recherche-Komponente definiert werden. Es

können sowohl übergreifende Zeiträume (Defaults) als auch spezifische Zeiträume je Suchprozess definiert werden.

Neue zu durchsuchende URLs werden automatisch aus den Internet-Suchergebnissen und den gecrawlten Inhalten generiert.

→ Teil A, Abschnitt 2.2 Recherche

01-008: Das Tool muss die Möglichkeit bieten, dass der Nutzer die Zeiträume der Durchsuchung manuell definieren kann

Für alle Recherche-Prozesse (automatisierte Suchen, Crawling) können die Zeiträume, auch periodische, manuelle über das Admin-UI der Recherche-Komponente definiert werden.

→ Teil A, Abschnitt 2.2 Recherche

01-009: Das Tool muss die Möglichkeit bieten, Trainingsdaten, Kategorien oder Rechtsverstöße hinzuzufügen

Trainingsdaten können über das Admin-UI des Systems direkt in das System importiert werden, auch unabhängig von zeitgesteuerten Prozessen. Kategorien, also neue Profile, können über das Admin-UI definiert werden. Vom Monitoring-Team festgesellte (bestätigte) Rechtsverstöße werden (semi)automatisch für die kontinuierliche Anpassung der Kategorien und ML-basierten Verfahren verwendet.

→ Teil A, Abschnitt 2.3.3 **Klassifikation und Identifikation**

→ Teil A, Abschnitt 2.3.5 **Integration externer Dienste**

→ Teil A, Abschnitt 2.3.6 **Abschließende Bewertung und Kategorisierung**

01-010: Das Tool muss die gefundenen potenziellen Rechtsverstöße in einer Datenbank abspeichern und hierarchisch abbilden können. 1.Anbieter / 2.Angebote / 3.Kategorien (Hate Speech, Jugendschutz relevante Inhalte, Schwerpunktanalyse)

Alle relevanten Informationen zu den Inhalten, die als potenziell rechtswidrig eingestuft wurden, werden strukturiert in einer Datenbank abgelegt. Hierfür wird die dokument-orientierte Datenbank verwendet.

→ Teil A, Abschnitt 2.4.1 **Datenbank der identifizierten Inhalte**

01-011: Das Tool muss die Möglichkeit bieten, dass Ergebnisse / Angebote der Text, Bild, Audio-Analysen priorisiert dargestellt werden

Für die abzuarbeitenden Inhalte werden in der Task-Management-Komponente Aufgaben angelegt. Jede Aufgabe erhält eine Priorität (z.B. Blocker, Critical, Major, Minor, Trivial), die regelbasiert aus verschiedenen Attributen des zugeordneten Inhalts (u.a. Reichweite, Menge³⁰, Kategorien, Wahrscheinlichkeit des Rechtsverstößes) abgeleitet wird.

In dem Workflow-UI können die Aufgaben nach sortiert, gefiltert oder auch Bearbeitern zugeordnet werden. Eine direkte Sortierung oder Filterung nach den einzelnen Attributen (z.B. Reichweite) ist unabhängig davon auch möglich.

Die initiale Konfiguration der Regel und auch mögliche Werte für Prioritäten sind in der Task-Management-Komponente konfigurierbar. Die Konfiguration erfolgt im Rahmen des Projekts in Abstimmung mit dem AG. Die Regel kann auch im Betrieb über das Admin-UI angepasst werden.

→ Teil A, Abschnitt 2.4.2 **Workflow zur Abarbeitung der identifizierten Inhalte**

³⁰ Eine Aufgabe kann sich auf eine zusammenhängende Menge von Inhalten beziehen, siehe auch Teil A, Abschnitt 2.4.2.

01-012: Das Tool muss die Möglichkeit bieten, nach Kategorien und Datenbankhierarchien (Anbieter, Reichweite...) zu filtern

Für den Zugriff auf die in der Datenbank gespeicherten Inhalte steht ein such-basiertes Web-UI zur Verfügung, mit dem die Inhalte insbesondere nach Kategorien, Anbietern und weiteren Merkmalen (z.B. Reichweite) gefiltert werden können.

Das Standard-Auswertungstool gestattet darüber hinaus auf einfache Art und Weise alle denkbaren Auswertungsmöglichkeiten.

→ Teil A, Abschnitt 2.4.1 **Datenbank der identifizierten Inhalte**

01-013: Das Tool muss einen eindeutigen Schlüssel/ID pro Beschwerde, Verstoß, Anbieter, Kategorie automatisiert vergeben. Es muss eine Dokumentation erfolgen. Die zur Dokumentation des Rechtsverstoßes notwendigen Daten müssen gespeichert werden

Die Aufgaben-Id in der Task-Management-Komponente fungiert als eindeutiger Schlüssel für den zugeordneten Verstoß.

Zur Dokumentation eines Rechtsverstoßes kann der Nutzer eigene Dateien (z.B. Screenshots) hochladen, als auch Dokumente aus in der Datenbank gespeicherten Analyseergebnisse automatisch generieren lassen, die dann entweder als Anhang mit der Aufgabe in der Task-Management-Komponente oder alternativ in der Datenbank gespeichert werden.

→ Teil A, Abschnitt 2.4.2 **Workflow zur Abarbeitung der identifizierten Inhalte**

01-014: Das Tool soll Basisstatistiken über Anbieter, Angebote / Verstöße anzeigen können

Alle Verstöße und auch Informationen zu identifizierten Anbietern und zu Angeboten werden strukturiert in einer Datenbank abgelegt. Das Standard-Auswertungstool gestattet auf einfache und flexible Art und Weise alle denkbaren Auswertungsmöglichkeiten.

→ Teil A, Abschnitt 2.4.1 **Datenbank der identifizierten Inhalte**

01-015: Die Daten, die durch das System erfasst werden, müssen durch eine Schnittstelle auch in einer zentralen Datenbank der LFM NRW speicherbar sein

Alle Daten zu Inhalten, identifizierten Anbietern und Angeboten werden strukturiert in einer Datenbank abgelegt. bietet einerseits verschiedene Möglichkeiten der direkten Abfrage von Inhalten über eine Standard-API. Insbesondere ist auch die Verwendung von SQL-Abfragen per CLI, REST-Endpunkt, ODBC oder JDBC möglich.

Andererseits können Inhalte auch mit standardisierten Mechanismen exportiert werden.

Eine automatisierte Anbindung (Export) einer zentralen Datenbank beim AG kann als Erweiterung implementiert werden.

01-016: Das System muss die Möglichkeit bieten, externe Quellen über eine API Schnittstelle anzubinden

Das System ist grundsätzlich offen für den Import von externen Daten. Evtl. automatisierte Importe sind individuell zu konzipieren und implementieren.

Ein Bulk-Importmechanismus für zu crawlende URLs steht sowohl in der Admin-UI der Recherche-Komponente als auch in der Workflow-UI zur Verfügung. Es können mehrere URLs mittels Copy & Paste über ein entsprechendes Eingabefeld eingegeben werden oder es kann eine CSV-Datei mit einer Liste von URLs hochgeladen werden.

01-017: Das System kann sowohl an relationale als auch nicht-relationale Datenbanken (z.B. Big Data Datenbanken) angebunden werden können

→ Siehe auch Antwort zu 01-015: Die Daten, die durch das System erfasst werden, müssen durch eine Schnittstelle auch in einer zentralen Datenbank der LFM NRW speicherbar sein.

01-018: Das System kann das Ergebnis eines externen Reporting Tools integrieren

Ergebnisse externe Reporting-Tools können auf verschiedene Weise in das System integriert werden, z.B. Import von Ergebnisdokumenten (PDF), Inline-Integration einer URL oder Integration einer externen Verlinkung.

7.2 Allgemeine und nicht-funktionale Anforderungen

02-001: Das Zielsystem muss eine Web-GUI zur Verfügung stellen

Das System stellt ein über alle Komponenten integriertes, web-basiertes UI zur Verfügung.

02-002: Das Zielsystem muss die Möglichkeit bieten, den Übertragungsstatus der Daten und weitere Loginformationen zu protokollieren

Alle wesentlichen Statusinformationen aller Komponenten des Systems werden in Log-Dateien protokolliert und in einem zentralen Log-Monitoring zusammengeführt. Bei Bedarf können spezifische Log-Parameter oder Events zusätzlich protokolliert werden.

02-003: Das Zielsystem sollte ergonomisch und benutzerfreundlich gestaltet sein

Bei der Entwicklung des Systems werden allgemein die Anforderungen an die Gebrauchstauglichkeit nach Norm EN ISO 9241 Teil 11, die Grundsätze der Dialoggestaltung

nach EN ISO 9241-110 und die Anforderungen an die Informationsdarstellung nach EN ISO 9241-12 berücksichtigt.

Im Rahmen des Projekts erfolgt eine Abstimmung der Benutzeroberfläche auf die spezifischen Aufgaben der Anwender des AG.

02-004: Das Zielsystem muss webbasiert zur Verfügung gestellt werden

Das bereitgestellte System stellt für die Anwender ein web-basiertes UI zur Verfügung.

02-005: Das Zielsystem muss die Browser Chrome, Firefox und Microsoft Edge unterstützen

Das bereitgestellte System unterstützt die Browser Chrome, Firefox und Edge in den zu Projektbeginn aktuellen Versionen.

02-006: Das Zielsystem muss in deutscher Sprache sein

Das UI des bereitgestellten Systems ist in deutscher Sprache.

02-007: Das Zielsystem muss mit mehreren Anwendern parallel nutzbar sein (unter Berücksichtigung der Performance)

Die initiale Konfiguration des bereitgestellten Systems wird so ausgelegt, dass es mit mindestens 20 Endanwendern performant nutzbar ist. Das System ist auch für eine größere Nutzeranzahl horizontal skalierbar. Die Architektur des Systems ermöglicht die Trennung potenziell aufwändiger Recherche- und Analyseprozesse von den interaktiven Komponenten, so dass die Performance für die Endanwender zu jeder Zeit sichergestellt werden kann.

02-008: Die Zielsysteme müssen DSGVO konform sein

Das bereitgestellte System erfüllt die Anforderungen der DSGVO an die Verarbeitung personenbezogener Daten – u.a. durch

- Zugriffskontrolle
- Vergabe von verschiedenen Berechtigungen
- Protokollierung
- https-Verschlüsselung der Kommunikation
- Datensicherung
- Export von Daten (Datenübertragbarkeit)

02-009: Das Zielsystem muss die Möglichkeit bieten, die Daten vor unbefugtem Zugriff Dritter zu schützen

Durch eine geeignete Sicherheitskonfiguration des Systems werden die Daten vor unbefugtem Zugriff Dritter geschützt. Dazu gehören u.a. die Verschlüsselung aller Verbindungen (d.h. Zugriff über https) und die Absicherung aller Zugriffe durch Authentifizierung.

02-010: Das Zielsystem muss immer sicherstellen, dass die Integrität der internen Datenbank nicht verletzt wird

Die Integrität der internen Datenbank wird gewährleistet durch

- die Implementierung wohldefinierter Schnittstellen für das ,
- eine geeignete Sicherheitskonfiguration des Systems und
- Backup & Recovery-Mechanismen.

02-011: Der Anbieter sollte bereit sein, dedizierte Vor-Ort-Schulungstermine zu organisieren

Condat ist gerne bereit, Vor-Ort-Schulungstermine zu organisieren.

→ Teil A, Abschnitt 3 **Vorgehensvorschlag und Projektplan**

02-012: Die LFM NRW muss die Möglichkeit haben, den Inhalt und Umfang der Schulungen mitbestimmen zu können

Inhalt und Umfang der Schulungen werden im Rahmen des Projekts mit dem AG individuell abgestimmt.

→ Teil A, Abschnitt 3 **Vorgehensvorschlag und Projektplan**

02-013: Der Anbieter sollte die Schulungsunterlagen und Kursmaterialien möglichst kundenindividuell zur Verfügung stellen

Schulungsunterlagen und Kursmaterialien werden im Rahmen des Projekts kundenindividuell zur Verfügung gestellt.

→ Teil A, Abschnitt 3 **Vorgehensvorschlag und Projektplan**

02-014: Initial muss es eine Präsenzschiilung zur Einführung des Systems in deutscher Sprache geben

Eine initiale Präsenzschiilung in deutscher Sprache ist Bestandteil des Angebotes.

→ Teil A, Abschnitt 3 **Vorgehensvorschlag und Projektplan**

02-015: Es muss eine Dokumentation / Handbuch in deutscher Sprache zur Verfügung stehen

Es werden eine deutschsprachige Dokumentation und Handbuch zur Verfügung gestellt:

- Benutzerhandbuch
- Fachliche Konfiguration und Administration
- Systemdokumentation

→ Teil A, Abschnitt 3 **Vorgehensvorschlag und Projektplan**