Dokumentation der Ermittlung des Vorgangs

Das Ziel dieser Dokumentation ist es, das entwickelte Analyseverfahren zur Zuordnung der Vorgangsarten eines Dokuments zu verstehen. Eine detaillierte Ausführung der Teilanalyseverfahren (Trefferanalyse, Hierarchieanalyse und Ausschlussanalyse) und der darin verwendeten Schlüsselwörter sowie die Anwendung verschiedener Werkzeuge würde den Rahmen dieser Dokumentation sprengen und wahrscheinlich mehr Fragen aufwerfen, als das Grundkonzept im Auge zu behalten. Falls dennoch Interesse besteht, können Fragen per E-Mail an [Memo.Zaninger@msg.group](mailto:Memo.Zaninger@msg.group) gestellt werden.

Die Ermittlung des Dokumententypen erfolgt durch die folgende Kategorisierung:

1. Genehmigung
2. Anhörung
3. Versagung
4. Bauverfahren
5. Nachforderung
6. Eingang
7. Anfrage
8. Antrag

Eine Übersicht über den Analyseprozess ist in Abbildung 1 zu finden.

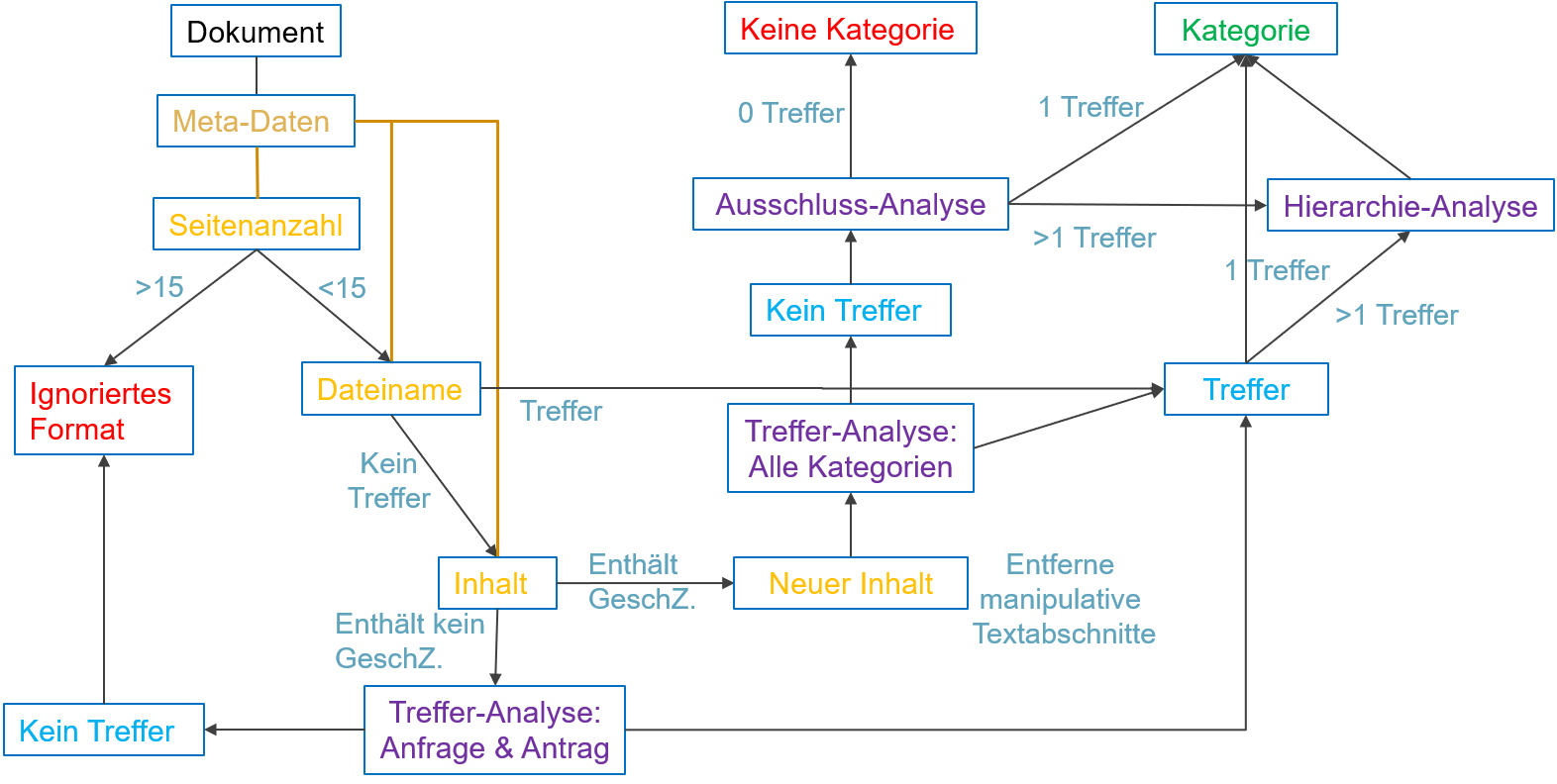


Abbildung : Prozessübersicht

Erklärung der Schritte:

1. Dokumente die eine größere Seitenanzahl als 15 aufweisen, sind fast nie als Behördendokumente zu klassifizieren. Zudem verringern sie die Geschwindigkeit der Analyse erheblich. Die betroffenen Dokumente repräsentieren Kataloge, Auszüge aus Büchern oder Veranstaltungsinformation und werden aus der Analyse gesondert entfernt.
2. Sofern der Dateiname einen Hinweis auf den Vorgangstypen gibt, so wird das Dokument entsprechend deklariert. Andernfalls wird der Inhalt der Dokumente herangezogen.
3. Der Inhalt wird auf den Verweis „GeschZ“, der für Geschäftszeichen steht, untersucht.   
   Ist das Zeichen vorhanden, so wird die Treffer-Analyse hinsichtlich aller Kategorien durchgeführt, nachdem manipulative Textabschnitte, wie bspw. Auflagen oder rechtliche Hinweise vom Inhalt des Dokuments entfernt wurden. Falls kein „GeschZ“ enthalten ist, so wird dieselbe Treffer-Analyse nur hinsichtlich der Kategorien Anfrage und Antrag durchgeführt, da solche Dokumente unter den aufgelisteten Kategorien diese vornehmlich enthalten. Hierbei ist ein Antrag als ein Antragsformular festgehalten, das exemplarisch der Abbildung 2 entspricht.
4. Falls die Treffer-Analyse zu einem eindeutigen Treffer führt, so ist die gesuchte Kategorie gefunden. Sofern mehr als ein Treffer vorliegt, so wird eine Hierarchieanalyse zur Eindeutigkeit eingesetzt.
5. Falls kein Treffer vorliegt, so kann immer noch die Ausschlussanalyse, zu einem Treffer führen.

Die genannten Analyse-Verfahren, werden nun folgend erläutert:

Treffer-Analyse:

Die Identifizierung einer Kategorie erfolgt über die Erkennung und unter Ausschluss von bestimmten Schlüsselwörtern.

Erkennung von Schlüsselwörtern:  
Die Erkennung der Kategorien wird im Wesentlichen über Schlüsselworte, die für jede Kategorie individuell sind, erreicht. Diese sind so allgemein wie möglich gehalten, um verschiedene Ausdrucksarten miteinzuschließen.

Die Kategorien „Genehmigung“, „Anhörung“ und „Versagung“ (1-3) werden durch die Identifizierung einer bestimmten Kombination von zwei verschiedenen Schlüsselworten bestimmt. Der Grund für diese Zwei-Faktor Identifizierung liegt an der Überschneidung von Schlüsselwörtern der jeweiligen Kategorien. So kann für der Begriff „Genehmigung“ in allen drei Kategorien für eine Übereinstimmung sorgen. Nur eine weitere Abstimmung durch bspw. „erteilt, angenommen, bestätigt, …“ oder „abgelehnt, nicht erteilt, nicht angenommen“ kann für eine Eindeutigkeit sorgen.   
Die Kategorien 4-8 können durch einen einzelnen Treffer mit einem Schlüsselwort erkannt werden, da diese eindeutiger und klarer voneinander abgetrennt sind.

Ausschluss durch Schlüsselwörter:

Auch wenn die Erkennung von Schlüsselworten zu einem Treffen führt, kann der Treffer nichtig werden, sobald bestimmte Schlüsselwörter im Schreiben auftauchen. Dies verhindert, dass trotz der Zwei-Faktor-Identifizierung, mehrere Treffer oder gar falsche Treffer vorliegen. Zudem kann diese Vorgehensweise zwei naheinanderliegende Kategorien separieren. Dies ist besonders für die Kategorien „Anhörung“ und „Versagung“ oder „Anfrage“/„Antrag“/„Nachforderung“ mit allen anderen Kategorien der Fall. So kann eine Anhörung den Begriff Versagung enthalten. Der umgekehrte Fall trifft selten zu (nur mit der Konstellation „Anhörung … vom …“). Falls in einem Schreiben die Begriffe „Versagung“ und „Anhörung“ ermittelt werden, so kann eine Anhörung ausgeschlossen werden (falls „Anhörung…vom“ nicht enthalten).

Hierarchie-Analyse:

Wie oben erwähnt existiert eine gewissen Hierarchie (wie zwischen Anhörung und Versagung) zwischen den Kategorien und ihren Schlüsselworten. Solch eine Hierarchie existiert auch zwischen einer „Nachforderung“ und einer anderen Kategorie (und auch in anderen Konstellationen). Eine Nachforderung bleibt eine Nachforderung (z.B. bei Forderung nach fehlenden Unterlagen), auch wenn bspw. eine Genehmigung oder Versagung in Aussicht gestellt wird. Folgende Hierarchie wird in diesem Rahmen verwendet:

1. Nachforderung, Eingang
2. Anfrage,
3. Antrag
4. Anhörung
5. Versagung
6. Genehmigung
7. Bauverfahren

In der Hierarchieanalyse wird die oben gelistete Hierarchie berücksichtigt, sofern mehrere Treffer vorliegen. In einem Rang können simultan mehrere Kategorien stehen, da diese sich hinsichtlich ihrer Schüsselwörter und des Inhalts des Dokuments nicht überschneiden oder sich nicht ausschließen.

Ausschluss-Analyse:

In der Ausschluss-Analyse werden die Hierarchie und die Ausschluss-Schlüsselwörter verwendet. In diesem Verfahren werden die Kategorien berücksichtigt, die definitiv ausgeschlossen werden. Wenn beispielsweise nur die Kategorie Genehmigung ausgeschlossen wird, so entspricht dies einer Versagung, da weder Anhörung noch andere Kategorien als Treffer erkannt wurden.

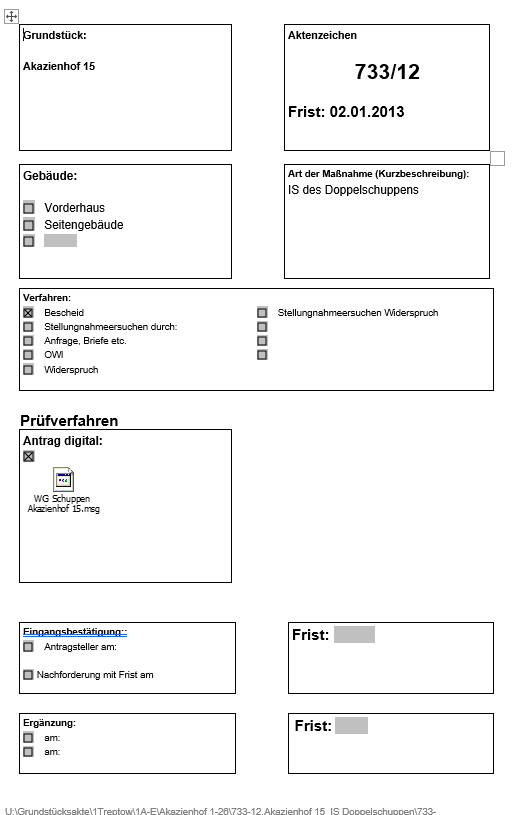


Abbildung : Zu sehen ist ein Antragsformular.