Konfiguration und Verwaltung

ELO iSearch

Inhaltsverzeichnis

ELO iSearch-Leitfaden 3

Was leistet die ELO iSearch? 3
Die ELO iSearch in der Praxis 4
Technische Beschreibung 32
Glossar 62

Indexaufbau 64

Indexaufbau 64

Indizierung im Hintergrund 65

Einleitung 65
Installation 66

ELO iSearch-Leitfaden

Was leistet die ELO iSearch?

Die ELO ECM Suite bildet die Basis für eine nutzenorientierte und effiziente Bereitstellung von Informationen: Relevante Daten werden auch bei größeren Mengen in kürzester Zeit gefunden. Komfortabel und schnell können die richtigen Treffer ausfindig gemacht werden.

Hierbei stehen verschiedene Suchfunktionen zur Verfügung, sodass Sie zielgerichtet auf benötigte Dokumente sowie auf Informationen aus integrierten Drittanwendungen, wie ERP-, CRM- oder E-Mail-Systeme, zugreifen können. Die ELO iSearch unterstützt Ihre Suchanfrage beispielsweise durch die Autovervollständigung von Suchbegriffen während der Eingabe, durch die Korrektur möglicher Tippfehler und durch die Suche nach ähnlichen oder verwandten Begriffen.

Zudem schlägt die ELO iSearch beispielsweise bei der Eingabe des Suchbegriffes `Vertrag` auch Synonyme wie `Vereinbarung` oder `Abkommen` vor. Und selbst wenn kein eindeutiger Suchbegriff definiert werden kann, führt die ELO iSearch zum Ergebnis. Es werden z. B. Begriffe zur Auswahl vorgeschlagen, die im Zusammenhang mit dem eingegebenen Suchwort stehen. Intelligente Wörterbücher sind darüber hinaus in der Lage, zusammengesetzte Begriffe, wie z. B. `Versicherungsvertrag` in seine Grundbestandteile zu zerlegen. Auch das Zurückführen von Suchbegriffen auf die jeweilige Grundform ist möglich. So müssen Sie sich nicht mehr um Pluralformen, gebeugte Wortformen oder um beliebig zusammengesetzte Begriffe bei der Formulierung Ihrer Suchanfrage kümmern. Dies erledigt die Suchtechnologie automatisch.

Bei Eingabe eines falsch geschriebenen Suchbegriffs, wie z. B. `Vertrahg`, erscheint zudem eine Meldung: `Meinten Sie vielleicht Vertrag?`. Diese intelligenten Assistenzsysteme fangen bereits frühzeitig Eingabefehler ab, damit wird das Suchergebnis verbessert oder oft erst grundsätzlich sichergestellt.

Bei der strukturierten Suche in den Metadaten sind die Suchanfragen häufig sehr präzise und gezielt. Die Suche nach einer Rechnungs-, Angebots- oder Vertragsnummer über das gesamte Repository funktioniert schnell und zuverlässig.

Eine weitere Herausforderung besteht darin, häufig verwendete, immer wiederkehrende oder komplexe Suchanfragen mit zum Teil aufwendig zusammengestellten Filteroptionen zu speichern und bei Bedarf schnell und komfortabel aufrufen zu können. Die bereits mit der ELO ECM Suite 10 eingeführten Kacheln zum Speichern und Verwalten von Suchfavoriten sind hierfür die perfekte Lösung.

Die ELO iSearch in der Praxis

Grundlagen und die Zwei-Wege-Strategie

ELO iSearch

Die ELO iSearch ermöglicht die Suche im Volltext, also dem Text innerhalb eines Dokuments und in den zusätzlichen Informationen wie den Metadaten. Sie bietet dabei verschiedene Möglichkeiten, Suchbereiche miteinander zu kombinieren. Der Volltext wird dabei analysiert und mit Suchregeln und Suchmustern (Algorithmen) verarbeitet.

Parallel zur SQL-Datenbank wird automatisch ein spezieller Suchindex aufgebaut und aktualisiert. So werden schnelle Antwortzeiten auch bei komplizierten Suchanfragen ermöglicht. Allerdings bedeutet dies auch eine leichte Verzögerung, bis neue Dokumente mit der ELO iSearch gefunden werden können. Üblicherweise sollte diese Verzögerung unter einer Minute liegen.

Die folgenden Abschnitte beziehen sich auf die Benutzung der ELO iSearch im ELO Java Client.

Metadatensuche

Die Metadatensuche ist eine zweite Suchform für spezielle Anforderungen. Sie sucht immer exakt nach den eingegebenen Werten direkt in der SQL-Datenbank und unterliegt daher nicht der Verzögerung.

Im Vergleich zur ELO iSearch kann es bei derselben Sucheingabe zu unterschiedlichen Ergebnissen kommen. Typischerweise wird dabei die ELO iSearch eher mehr Treffer finden als die Metadatensuche, da die technologische Umsetzung der sprachspezifischen Suchregeln und Suchmuster z. B. auch die Mehrzahlform eines Suchwortes findet.

Grundsätzlich ist die ELO iSearch der modernere, flexiblere und leistungsfähigere Ansatz und die bevorzugte Suchvariante in ELO.

Volltextsuche

Der Volltext bezeichnet alle Textinhalte eines Dokuments. Bei einer Rechnung wären dies beispielsweise neben Namen, Anschrift und Adresse des Empfängers auch die des Erstellers, die Rechnungspositionen, Beträge und auch eventuell vorhandene Zahlungshinweise.

Linguistik

Die Linguistik (Sprachwissenschaft) befasst sich mit den Bedeutungen und Regeln der menschlichen Sprache. Auf Basis solcher Regeln verwendet die ELO iSearch automatisierte Methoden, um Texte so aufzubereiten, dass sie möglichst einfach und komfortabel gefunden werden können. Dadurch lassen sich mit einfachen Suchworten auch Dokumente finden, die inhaltlich passen. Es ist also nicht nötig, die exakte Schreibweise im Dokument zu kennen. Dies ist vor allem immer dann praktisch, wenn es unterschiedliche Schreibvarianten gibt. Nachfolgend wird in diesem Zusammenhang von Suchmustern oder Sprach- und Suchregeln gesprochen.

Zerlegung an speziellen Zeichen

Bestimmte Zeichen haben für die textuelle Erfassung keine Bedeutung oder behindern diese sogar. Insbesondere Punktuationszeichen und Sonderzeichen am Anfang und am Ende eines Wortes werden deshalb entfernt. Ebenfalls entfernt werden Punktuations- und Sonderzeichen in der Mitte von Wörtern und Begriffen. Diese Begriffe werden dadurch in Einzelteile zerlegt. Dabei wird allerdings der Originalbegriff zusätzlich beibehalten.

Beispiel 1

Der Begriff `Holz-Haus` wird indiziert als `holz-haus`, `holz` und `haus`.

Auch an Stellen, wo innerhalb eines Wortes ein Großbuchstabe auf einen Kleinbuchstaben folgt (sogenannte CamelCase-Stellen), findet eine Auftrennung statt. Solche Begriffe kommen häufig in künstlich gebildeten Eigennamen und Programmiersprachen vor.

Beispiel 2

Der Begriff `MegaPowerShot` wird in die Teile `megapowershot`, `mega`, `power` und `shot` zerlegt.

Ein Begriff mit CamelCase-Stelle wird also genau so behandelt wie ein Begriff mit Bindestrich. Bei der Suche nach solchen Begriffen ist zu empfehlen, nach den Einzelteilen oder einer Kombination von Einzelteilen zu suchen, insbesondere, wenn es sich um längere zusammengesetzte Begriffe handelt und die Schreibweise nicht genau bekannt ist.

Kompositazerlegung

Zusammengesetzte Wörter werden als Komposita bezeichnet. Gerade die deutsche Sprache verwendet davon sehr viele, z. B. `Eingangsrechnung` oder auch der berühmte `Donaudampfschifffahrtskapitän`. Diese werden in ihre Einzelteile zerlegt, also z. B. `Eingang` und `Rechnung`.

Stammformbildung

Bei dieser Bearbeitung wird die Einzahl aus Mehrzahlformen gebildet oder der Infinitiv zu deklinierten Verben gefunden.

Beispiel 1

Zu dem Wort `Rechnungen` wird die Einzahl `Rechnung` ergänzt, zu `bezahlt` die Stammform `zahlen`. Umgekehrt ist die Bildung von Mehrzahlformen aus dem Singular eines Wortes jedoch nicht möglich. Dies wird unter anderem in der nachfolgenden Tabelle verdeutlicht und sollte bei der Auswahl und der Schreibweise des Suchbegriffes berücksichtigt werden.

> **Beachten Sie**
>
> Kompositazerlegung und Stammformbildung finden bei der Indexierung statt. Um dieses Feature zu nutzen, müssen Sie bei der Suche die Stammform eingeben. Verwenden Sie als Suchwort eine gebeugte Form, wird nach exakt dieser Form als einzelnes Wort gesucht. Beispiele hierzu finden Sie in der folgenden Tabelle.

Volltext	Suchwort	Treffer
Rechnungen	Rechnungen	Ja
Rechnungen	Rechnung	Ja
Rechnung	Rechnungen	Nein
Eingangsrechnungen	Eingangsrechnungen	Ja
Eingangsrechnungen	Rechnungen	Nein
Eingangsrechnungen	Rechnung	Ja
Rechnung	Rechnung	Ja

Beispiel 2

Die Suche mithilfe der iSearch unterstützt keine zusammengesetzten Teilwörter als Suchstring. Das Wort `Fachwerkhäuser` ist z. B. aus Sicht des Analyse-Plug-ins ein zusammengesetztes Wort, das in Token (Einzelbestandteile) zerlegt und durch Stammformbildung (`Häuser` wird zu `Haus`) weiter verändert wird. Die Zerlegung ergibt dabei als Token immer die kleinstmöglichen Teilwörter, aber keine zusammengesetzten Wörter wie `Fachwerk`. Im Index der Elasticsearch existieren dadurch nach der Analyse die folgenden Einzelwörter: `fach`, `werk`, `haus` und `fachwerkhaeuser` (als Original nicht zerlegt).

Bei einer Volltextsuche nach Dokumenten, die das Wort `Fachwerkhäuser` enthalten, ergeben sich z. B. folgende Suchmöglichkeiten:

Suchwort	Treffer	Kommentar
fachwerkhäuser	Ja	
fachwerkhaeuser	Ja	
fachwerkhaus	Nein	
fachwerk	Nein	
fachwerke	Nein	
fach	Ja	führt ggf. zu vielen nicht gewünschten Treffern
werk	Ja	führt ggf. zu vielen nicht gewünschten Treffern
haus	Ja	führt ggf. zu vielen nicht gewünschten Treffern
fach werk	Ja	Suchkombination mit UND-Verknüpfung
werk haus	Ja	Suchkombination mit UND-Verknüpfung
fach haus	Ja	Suchkombination mit UND-Verknüpfung

Sprachen

Grundsätzlich müssen Sie beachten, dass alle Suchregeln abhängig von der verwendeten Sprache gelten. Im Englischen wird eine Mehrzahl in der Regel durch ein angehängtes -s gebildet (z. B. `book` und `books`), im Deutschen gibt es hingegen verschiedene Varianten und auch viele unregelmäßige Formen (aus einem `Buch` werden mehrere `Bücher`).

Die ELO iSearch verwendet je nach Sprache unterschiedliche Regeln. Es stehen somit nicht in jeder Sprache Suchregeln und -muster im gleichen Umfang zur Verfügung. Die meisten Optimierungen der Suchregeln und -muster beziehen sich auf die deutsche Sprache. Die englische Sprache wird aber ebenfalls sehr gut unterstützt.

Um die Dokumente sinnvoll automatisch bearbeiten zu können, wird in einem ersten Schritt die Sprache des Volltextes anhand der verwendeten Wörter ermittelt. Diese wird dann zur Suchregel- und Suchmusterbearbeitung des Textes verwendet. Probleme kann es daher bei Dokumenten geben, die unterschiedliche Sprachen verwenden.

Die Felder der Metadaten werden immer mit der eingestellten/ausgewählten Sprache des Repositorys bearbeitet.

Mehrwortsuche

Geben Sie mehrere Wörter in das Suchfeld ein, wird nach Dokumenten gesucht, in denen alle diese Wörter vorkommen. Sie schränken also die Suche mit jedem zusätzlichen Wort weiter ein und finden im Idealfall genau das passende Dokument.

Um mehrere alternative Bezeichnungen zu suchen, trennen Sie diese mit einem Komma, z. B. `Dienstreise, Kundentermin`. Es genügt dann, wenn eines der Suchwörter im Dokument vorkommt.

> **Beachten Sie**
>
> Damit die Mehrwortsuche funktionieren kann, müssen alle gesuchten Wörter in einem Bereich vorhanden sein (siehe Suchbereich/Durchsuchte Felder). Eine Suche nach `Firmenfest Tischplan` in allen Suchbereichen findet also Dokumente, bei denen die Wörter `Firmenfest` und `Tischplan` entweder beide im Volltext vorkommen, beide in den Metadaten stehen oder beide im Feed genannt werden. Sie würden keine Dokumente finden, in denen die Wörter über mehrere Bereiche verteilt sind, z. B. das eine Wort im Volltext und das andere in den Metadaten.

Verwendung von Wildcards im Suchbegriff

Das Verfahren einer indexbasierten Suche basiert auf der Zerlegung in sogenannte Token zum Zeitpunkt der Indizierung und auf dem Token-Vergleich zum Zeitpunkt der Suche. Dadurch ist diese Suche weniger geeignet für die Verwendung von Wildcards (z. B. *), da dann eine zeichenbasierte Expansion zum Zeitpunkt der Suche durchgeführt werden muss.

Ein Suchbegriff mit * kann nicht mithilfe der linguistischen Mittel in Einzelteile zerlegt werden und die Suche ist in der Regel nicht zielgenau. Auch führt die Token-Expansion zum Suchzeitpunkt zu einer erheblichen Verringerung der Such-Performance.

Generell ist zu empfehlen, dass Wildcards in der ELO iSearch nur ausnahmsweise verwendet werden sollten, wenn man schnell ein beliebiges Ergebnis braucht. Besser ist die Suche mit Kombination von Einzelteilen (wahlweise UND-verknüpft mit Leerzeichen getrennt oder ODER-verknüpft mit Komma getrennt). Beachten Sie hierzu auch die Beispiele im Abschnitt Stammformbildung.

Phrasensuche

Es können nicht nur einzelne Wörter im Dokument gesucht werden, sondern auch längere Begriffe und ganze Sätze. Setzen Sie hierzu die zusammenhängenden Wörter in Anführungsstriche. Diese müssen dann genau in dieser Reihenfolge direkt hintereinander im Volltext vorkommen.

Mit einer Suche nach `Heinrich Müller` finden Sie genau diesen Namen. Ohne die Anführungszeichen würden Sie auch Dokumente finden, in denen die Wörter weiter auseinanderstehen, z. B. `Teilnehmer waren Heinrich Schulz und Dieter Müller`.

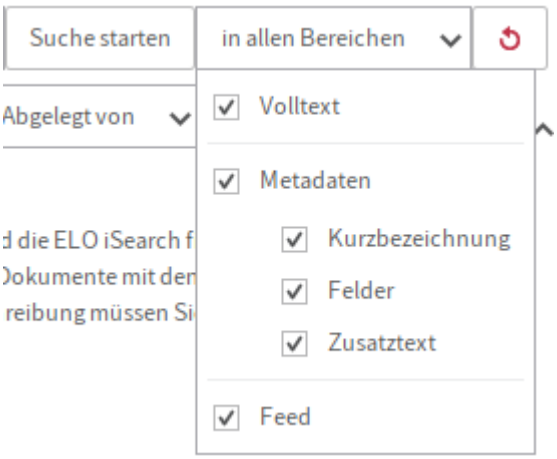
Thesaurus

In die ELO iSearch ist ein Thesaurus integriert, der Synonyme zu dem eingegebenen Suchwort auflistet. Diese erscheinen, sobald die Eingabe eines Wortes mit einem Leerzeichen beendet wird.

Wählen Sie ein Synonym aus, um es als Alternative in die Sucheingabe zu übernehmen. Der Thesaurus kann firmenspezifisch erweitert werden. Ebenso ist es möglich, Vorschläge zu entfernen. Lesen Sie dazu den Abschnitt Einbinden eigener Thesauri.

Suchbereich/Durchsuchte Felder

Direkt neben dem Eingabefeld und dem Button *Suche starten* befindet sich die Auswahl des Suchbereichs. Hiermit können Sie einstellen, wo nach den eingegebenen Wörtern gesucht werden soll. Zur Auswahl stehen die drei Bereiche *Volltext*, *Metadaten* und *Feed*. Der Bereich *Metadaten* ist in *Kurzbezeichnung*, *Felder* und *Zusatztext* unterteilt. Die Standardeinstellung ist die Suche in allen Bereichen.



> **Information**
>
> Es wird wie oben für die Volltextsuche beschrieben in allen Feldern mit den gleichen Suchregeln und -mustern gesucht. Stammformen werden erkannt oder Wortzusammensetzungen getrennt gesucht.

Volltext

Bei rein digitalen Dokumenten entspricht der Volltext immer genau dem Text im Dokument. Die Suche im Volltext kann also alle Inhalte des Dokuments finden. Im Gegensatz zu der Suche in Metadaten eignet sich die Volltextsuche dafür, nicht nur definierte Schlagworte oder Metainformationen zu finden, sondern alle Begriffe des Dokuments.

Gescannte Dokumente müssen erst mittels OCR in eine Textform umgewandelt werden. Dabei kann es immer zu Erkennungsfehlern einzelner Buchstaben oder ganzer Wörter kommen, welche bei der Suche das Finden des Dokuments verhindern. Zudem kann es durch Textblöcke und Spaltensatz zu einer Vermischung von Sätzen kommen, wodurch eine Phrasensuche den Satz nicht finden kann.

Kurzbezeichnung

Dokumente und Ordner werden in ELO mit ihrer Kurzbezeichnung angezeigt. Verwenden Sie daher eine möglichst treffende und korrekte Bezeichnung des Inhalts in der Kurzbezeichnung.

Felder

Mit der Einstellung *Felder* werden alle konfigurierten Felder in allen Masken in die Suche einbezogen. Suchen Sie in den Feldern nach dem Wort `Müller`, erhalten Sie beispielsweise folgende Treffer:

- `Müller Baustoffe GmbH` in einem Feld *Lieferant*
- `Heinrich Müller` in einem Feld *Kunde*
- `Müllerstraße 42` in einem Feld *Anschrift*

Zusatztext

Der Zusatztext (eine Registerkarte in der Maske) dient in ELO zur Aufnahme weiterer Informationen zu einem Dokument oder Ordner. Er bietet im Vergleich zu einem Feld Platz für längere Texte und kann hier ebenfalls durchsucht werden.

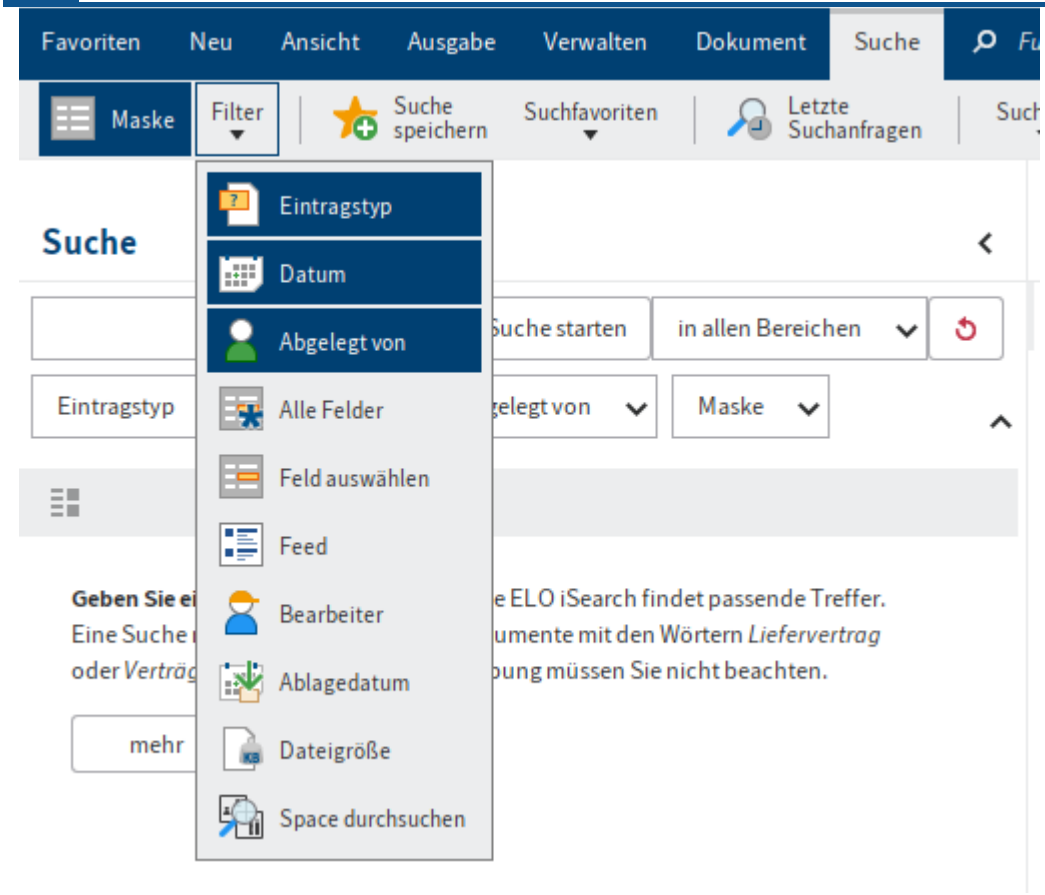
Feed

Mit der Einstellung *Feed* werden alle Beiträge und Kommentare im Feed in Textform durchsucht.

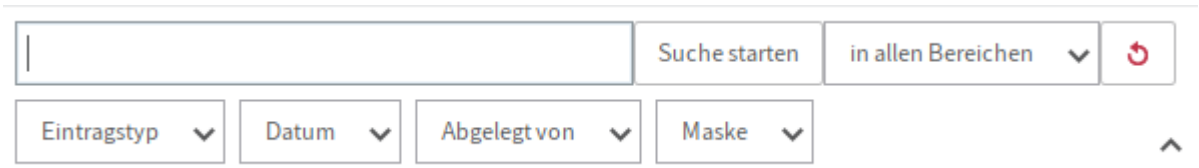
> **Information**
>
> Im Feed sind auch die Versionsnummern und Versionskommentare enthalten, sowohl zu der aktuellen Arbeitsversionen eines Dokuments als auch zu älteren Versionen.

Filter

In vielen Fällen ist es sinnvoll, die reine Textsuche um strukturierte Informationen zur Einschränkung des Suchbereichs zu ergänzen. Die ELO iSearch bietet hierfür Filter an. Diese können über das Menüband eingeblendet werden und erscheinen dann als zusätzliche Buttons unter dem Eingabefeld.



Eine Reihe von Filtern wird standardmäßig unterhalb des Suchfeldes angezeigt. Es handelt sich dabei um die Filter, die am häufigsten verwendet werden. Welche Filter das sind, können Sie sowohl global als auch gruppenspezifisch und individuell einstellen, siehe Abschnitt zu den Suchfavoriten.

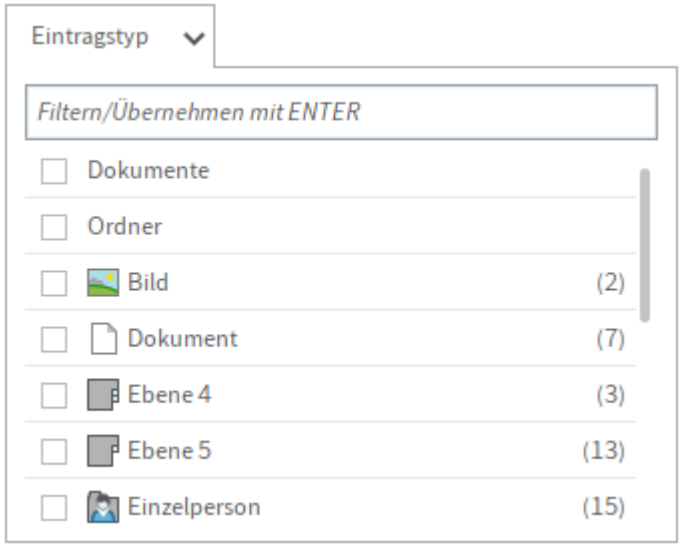


Durch Auswahl des Buttons erscheint der jeweilige Filter und Sie können spezielle Suchangaben vornehmen. Dabei ist fast immer eine Eingabe sowohl per Maus als auch per Tastatur möglich.

Die Filter berücksichtigen genauso wie die Suche die ELO Berechtigungsstruktur. Es werden also nur die Filterwerte zu den Einträgen angezeigt, zu denen eine Person auch berechtigt ist.

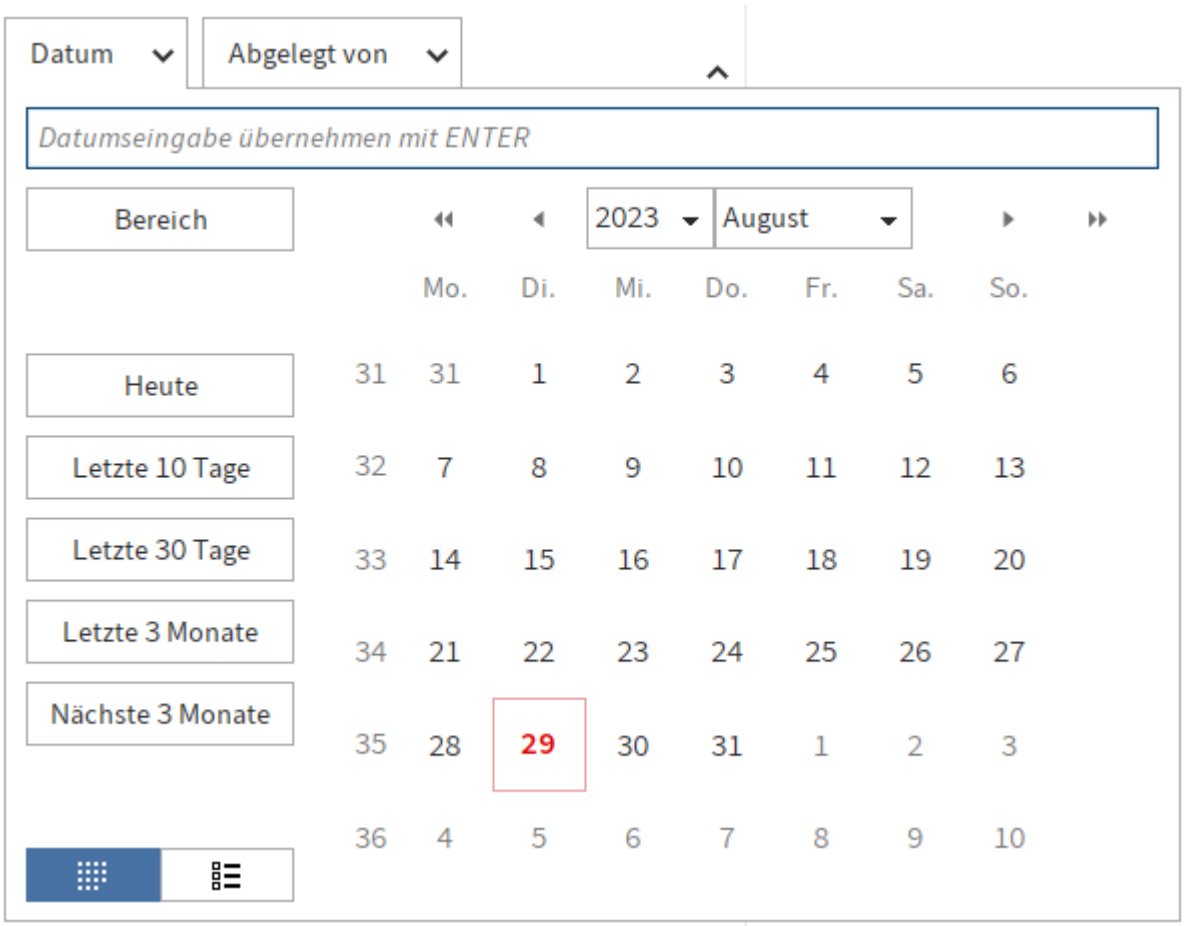
Eintragstyp

Der Filter für den Eintragstyp zeigt, wie der Filter für die Maske, eine Liste der möglichen Werte an. Die Bedienung ist ebenfalls dieselbe, mit Search-as-you-type und Autovervollständigung oder Auswahl aus der Liste. Als Besonderheit werden hier ganz oben die zwei Sammelbegriffe *Dokumente* und *Ordner* angeboten, welche alle entsprechenden Eintragstypen zusammenfassen.



Datum/Ablagedatum

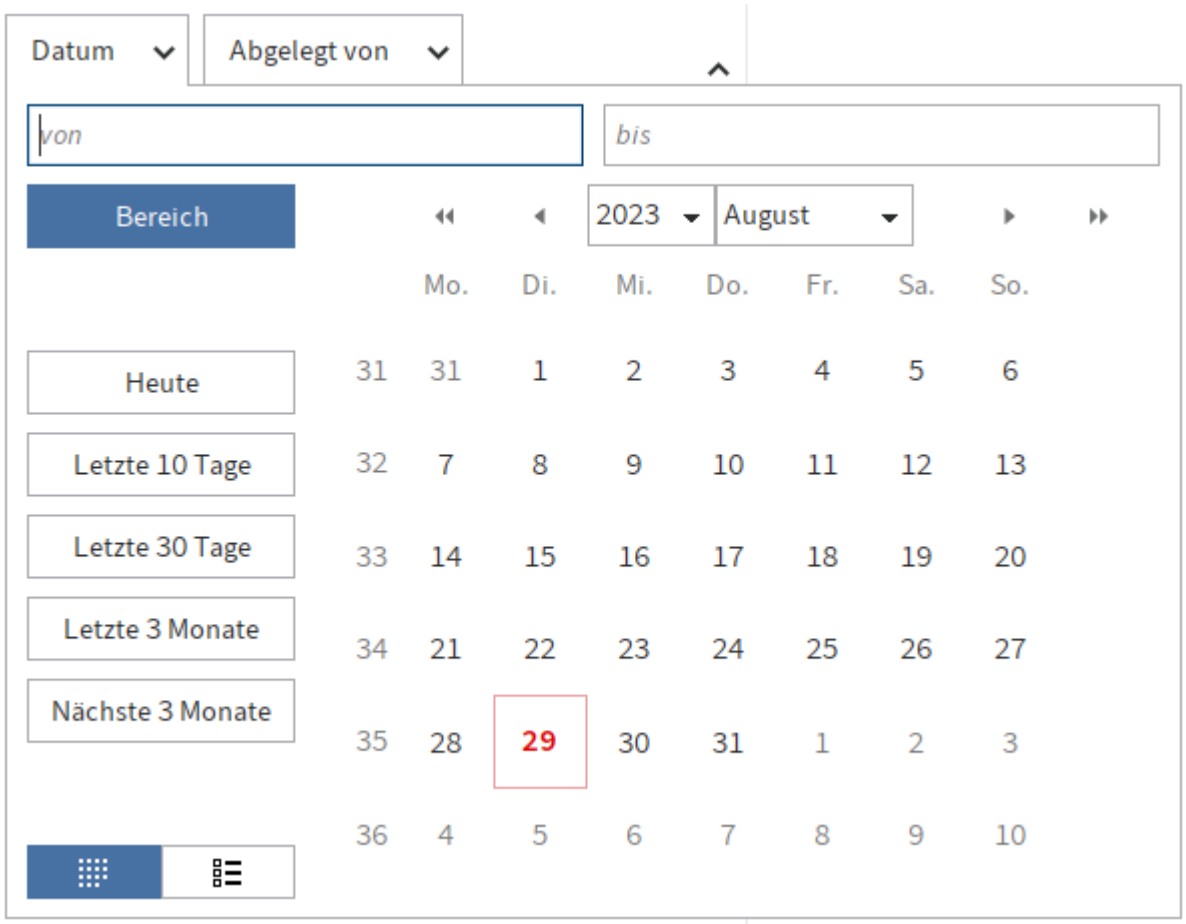
Der Filter für die Datumsangaben der Metadaten bietet viele Möglichkeiten zur Suche nach einem Datum. Es wird dabei immer nach ganzen Tagen gesucht, also von `00:00:00` Uhr bis `23:59:59` eines Tages.



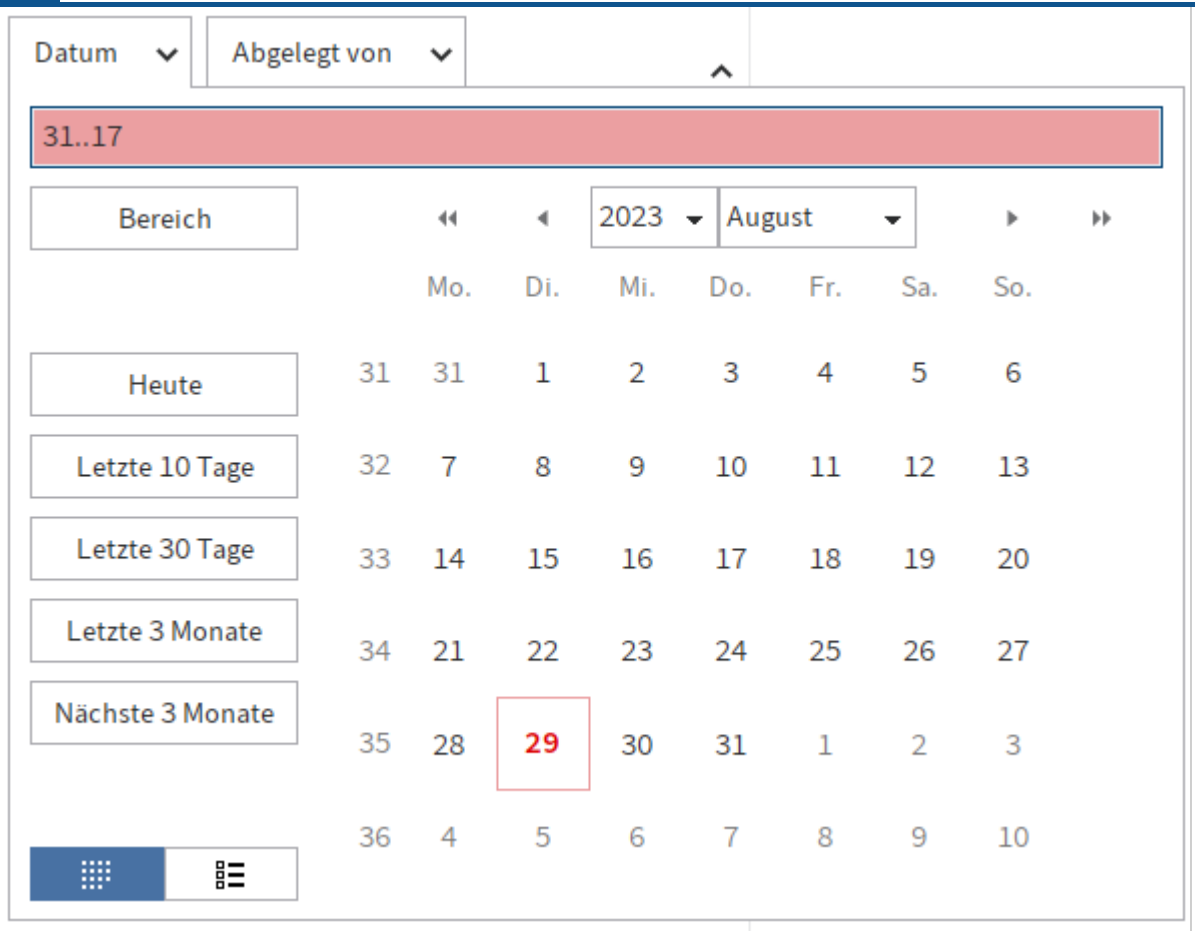
Ganz oben in der Abbildung finden Sie ein Eingabefeld, in das ein exaktes Datum eingegeben werden kann. Hierbei verwendet der ELO Java Client die in der Konfiguration eingestellten Datumsformate. Einen Zeitraum geben Sie mit drei Punkten als Trenner in der Mitte an, z. B.

`01.05.2020…10.05.2020`. Die Eingabe müssen Sie stets mit der EINGABE-Taste abschließen. Der Filter schließt sich und der Zeitraum wird in einem kleinen grauen Kasten dargestellt.

Um den Zeitraum jedoch noch komfortabler einzugeben, können Sie alternativ auf den Modus *Bereich* mit zwei Eingabefeldern wechseln. Nach der Eingabe des Wertes `von` springt der Fokus automatisch in das Feld `bis` und Sie können beide Felder auch mit einer Auswahl über den Kalender befüllen.



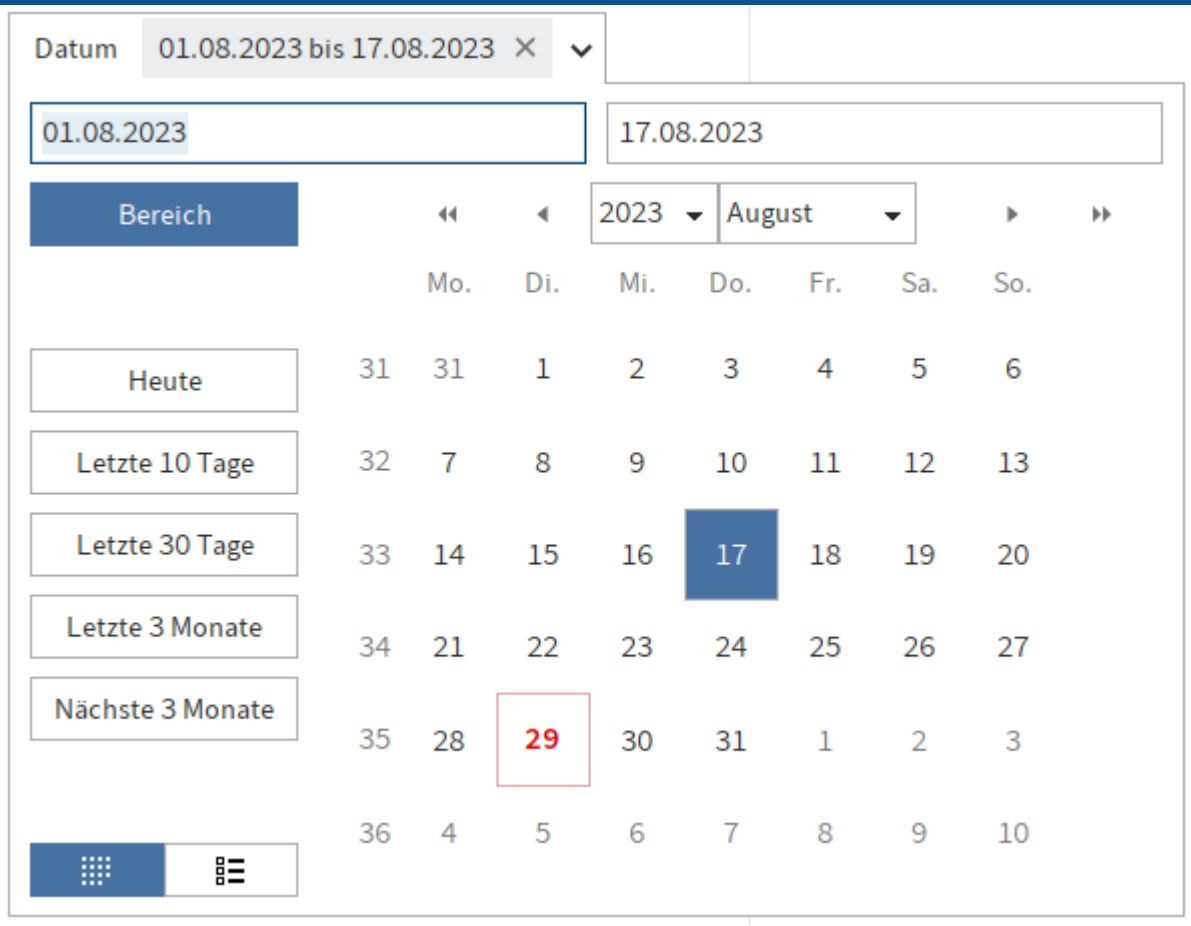
Bei einer nicht als Datum erkennbaren Eingabe bleibt der Filter nach der Betätigung der EINGABE-Taste geöffnet und das Eingabefeld wird rot eingefärbt, um den Fehler anzuzeigen.



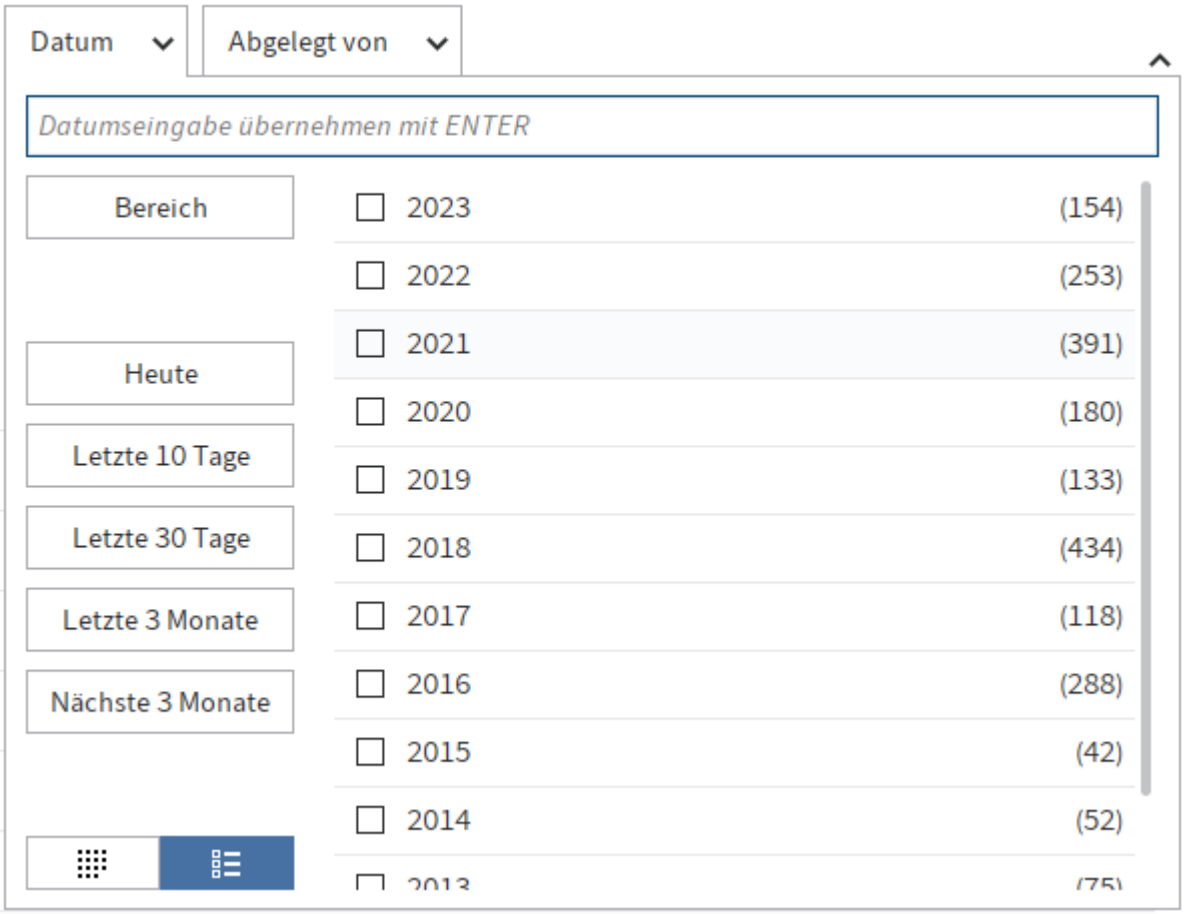
Es können auch relative Zeiträume eingegeben werden. Dazu verwenden Sie Minus und Plus als Vorzeichen für einen Zeitraum in Tagen, ausgehend vom aktuellen Datum. Die Eingabe `-2` sucht alles von vorgestern um 0 Uhr bis heute Ende des Tages. Mit einem angehängten `m` können Monate verwendet werden. `+1m` sucht alles von heute 0 Uhr bis einen Monat in der Zukunft (28, 29, 30 oder 31 Tage, je nach Monat und Jahr).

Links befinden sich fünf Buttons, die häufige Zeiträume direkt anbieten. Der Zeitraum wird nach der Auswahl auch in das Eingabefeld eingetragen, um die verwendete Syntax anzuzeigen.

Rechts wird ein Kalender zur Datumsauswahl angezeigt. Das aktuelle Datum ist rot hervorgehoben. Über die Pfeile über dem Kalender kann in Monaten und Jahren vor- und zurückgesprungen werden. Über das Pull-down-Menü im Jahresfeld bzw. Monatsfeld kann direkt der gewünschte Monat bzw. das gewünschte Jahr angewählt werden. Sie wählen per Mausklick einen einzelnen Tag aus. Mit Mausklick auf einen zweiten Tag wählen Sie den dazwischenliegenden Zeitraum aus.

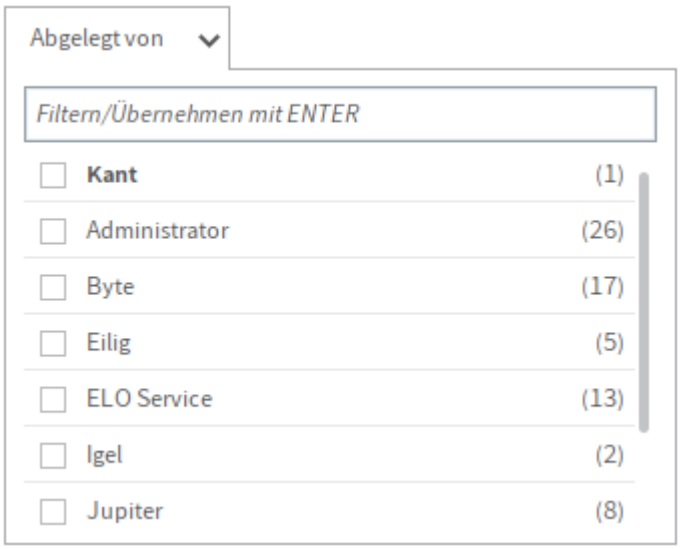


Links unten befinden sich zwei weitere Buttons, um zwischen dem Kalender und der Jahresübersicht in Listenform umzuschalten. Die Listenform zeigt alle vorhandenen Jahreswerte und jeweils rechts in Klammern die Anzahl der dazugehörigen Einträge. Sie können ein einzelnes Jahr (durch Auswahl eines Listeneintrags) oder mehrere Jahre (per Auswahl der Checkboxen) auswählen.



Abgelegt von/Bearbeiter

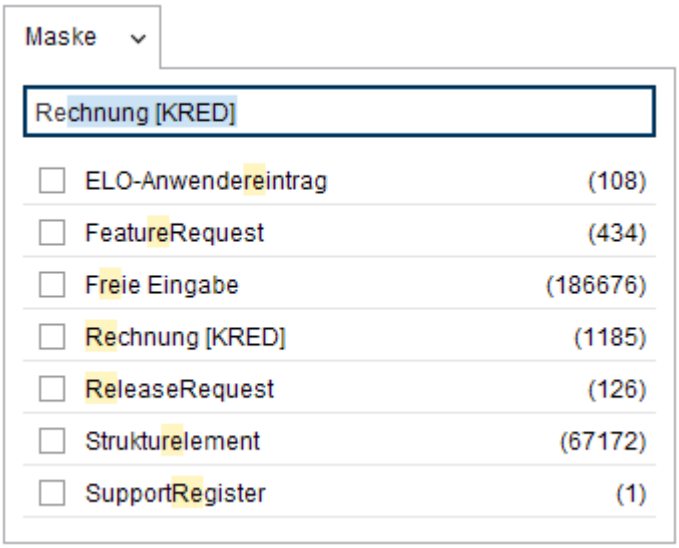
Die Filter für *Abgelegt von* und *Bearbeiter* gleichen denen von Maske und Eintragstyp. Die aktuell in ELO angemeldete Person wird hier ganz oben angezeigt.



Maske

Der Filter zeigt alle verwendeten Masken und jeweils rechts in Klammern die Anzahl der dazugehörigen Einträge. Die Masken sind in alphabetischer Reihenfolge sortiert. Sie können eine einzelne Maske oder mehrere Masken über die Checkboxen auswählen.

Durch Eingabe in das Filterfeld können Sie *Search-as-you-type* verwenden. Dabei wird Ihre aktuelle Eingabe sofort mit den möglichen Werten verglichen. Die Liste wird dadurch so eingegrenzt, dass nur noch die Masken angezeigt werden, die den eingegebenen Text enthalten. Es spielt keine Rolle, an welcher Stelle der Text steht. Im folgenden Beispiel wird daher durch Eingabe von `Re` sowohl `Rechnung` als auch `SupportRegister` gefunden. Die Trefferstelle wird hellgelb hinterlegt.



Bei der Eingabe wird außerdem eine Autovervollständigung vorgeschlagen. Dazu wird die Eingabe als Anfang des Maskennamens verstanden und die passende Maske mit den meisten Einträgen als weiterer Text in das Eingabefeld eingetragen und markiert. Mit der EINGABE-Taste können Sie die Autovervollständigung akzeptieren und übernehmen. Oder Sie tippen weiter, wenn Sie eine andere Maske suchen.

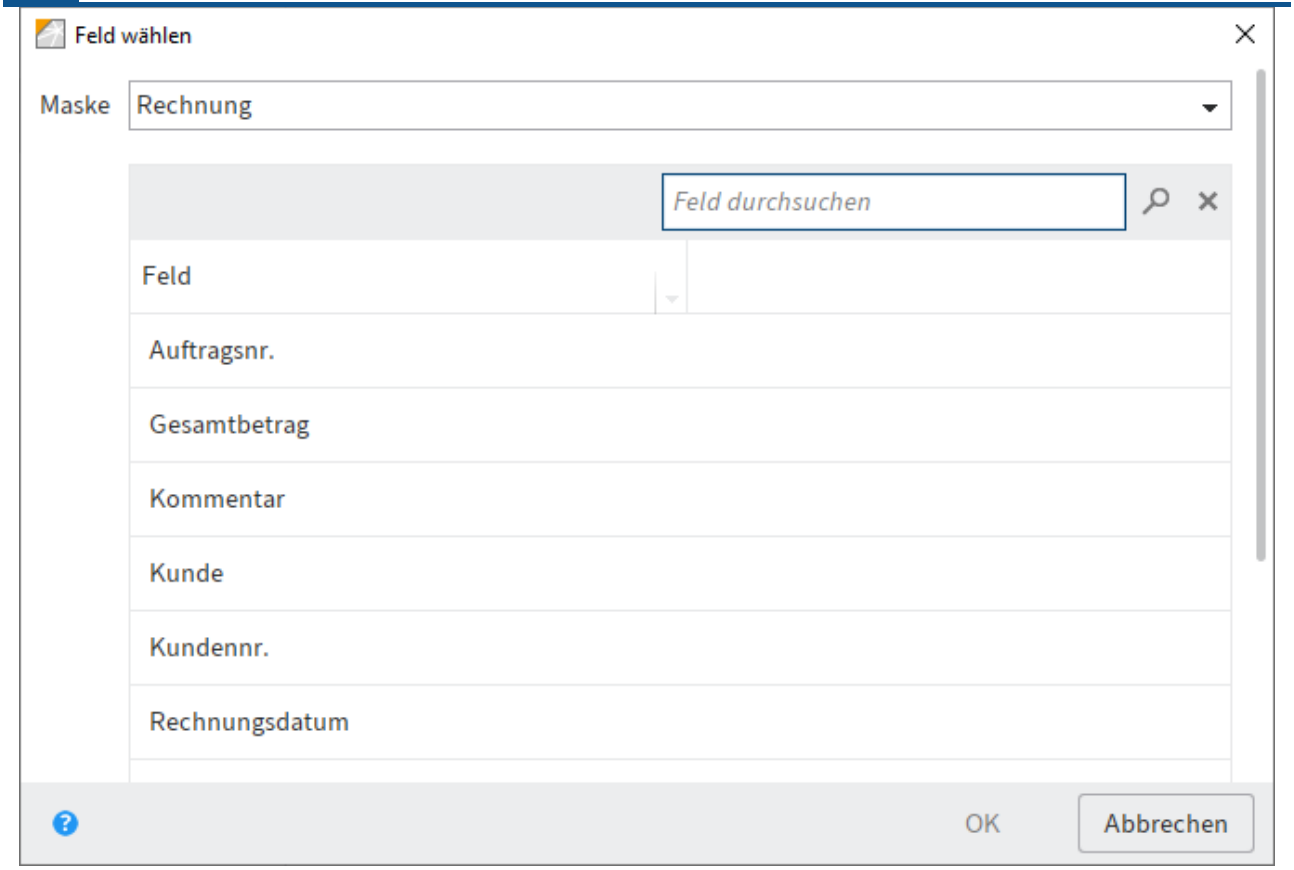
Alle Felder

Dieser Filter durchsucht alle konfigurierten Felder in allen Masken. Er verwendet dabei die gleichen linguistischen Regeln wie das Eingabefeld der iSearch und stellt einen einfachen Weg dar, um zwischen der Suche im Volltext (über das Eingabefeld) und in den Metadaten (mit diesem Filter) gezielt zu unterscheiden.

Wenn Sie zusammenhängende Begriffe, numerische Werte oder Datumsangaben finden möchten, ist die Suche in einem spezifischen Feld besser geeignet.

Feld auswählen

Filter für bestimmte Felder können Sie manuell hinzufügen. Dazu wählen Sie im Dialog zuerst die jeweilige Maske und dann die Felder aus. Wenn Sie die Suche bereits mit einem Maskenfilter auf eine Maske begrenzt haben, ist diese im Dialog vorausgewählt.



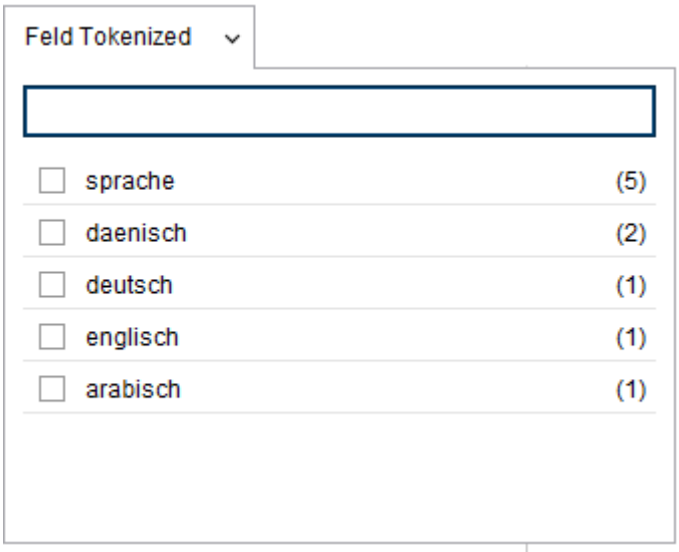
Information

Die Auswahl der Felder erfolgt für eine einfache Benutzung über die Maske. Der Feld-Filter an sich ist aber maskenübergreifend, wenn die Feldvorlage des ausgewählten Felds auch in anderen Masken verwendet wird. Um die Suche auf eine Maske einzugrenzen, ist ein Maskenfilter notwendig.

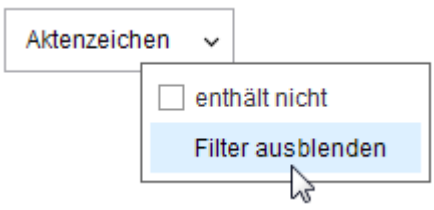
Der Filter für ein Feld kann mit Suchregel- und Suchmuster-Methoden oder kompletten Werten arbeiten. Dies ist eine Konfiguration der Feldvorlage.

Die Option *Tokenisierung deaktivieren* (siehe Abschnitt Tokenisierung) einer Feldvorlage schaltet die Suchregeln und -muster für dieses Feld aus. Es erscheinen in der Liste des Filters die exakten Werte der Metadaten. Die Suche arbeitet für dieses Feld dann ebenfalls in einem exakten Modus. Teilbegriffe werden in diesem Fall nicht berücksichtigt.

Bei aktiven Suchregeln und -mustern erscheinen in der Liste des Filters nur einzelne Wörter, sowohl die Originale als auch die erzeugten Stammformen. Die Suchmöglichkeiten in diesem Filter entsprechen dann denen des Eingabefeldes der Suche.

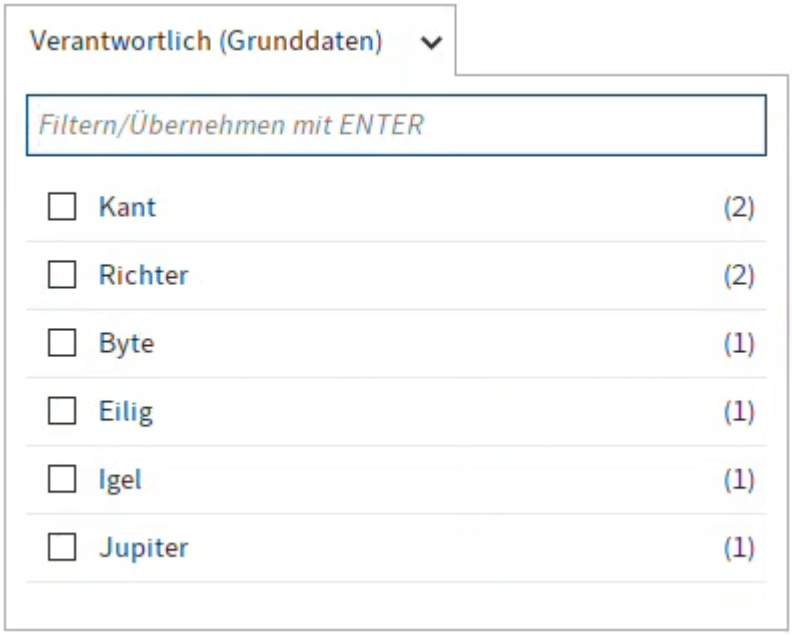


Einzelne Filter können Sie auch über das Kontextmenü ausblenden.



Benutzerfelder

Bei Benutzerfeldern zeigt die Liste des Filters die Display-Namen aller in dem Feld vorkommenden Benutzer als Suchvorschläge an.

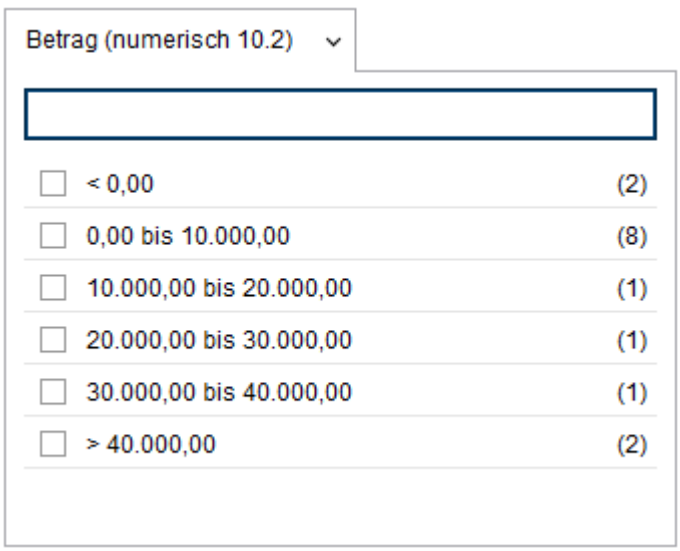


Beachten Sie

Die Feldinhalte eines Felds vom Typ *Benutzer* sind nur über die GUID des Benutzers mit dem ELO Benutzereintrag in der Datenbank verknüpft. Eine alleinige Änderung des Namens führt somit nicht zu einer Aktualisierung der Display-Namenseinträge in benutzerdefinierten Feldern der Maske. Deshalb können nach einer Namensänderung noch bisherige Werte im iSearch-Index vorhanden sein, solange bis durch eine anderweitige Änderung ein Update des betreffenden Dokuments getriggert wird. Eine automatische Triggerung wird aus Performance-Gründen nicht durchgeführt.

Numerische Felder

Bei Feldern, die als numerische Daten konfiguriert sind, zeigt die Liste des Filters Wertebereiche als Suchvorschläge an.

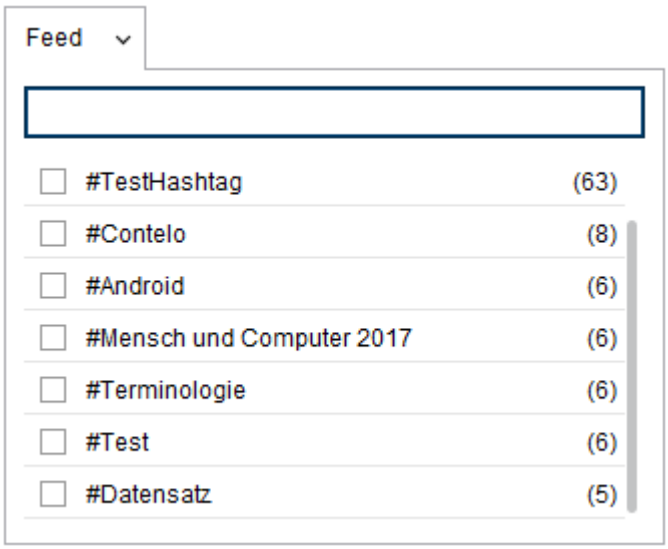


Solche Bereiche können auch direkt mit drei Punkten als Bereichstrennzeichen in den Filter eingegeben werden, z. B. `100...900` für eine Suche von 100 bis 900. Außerdem ist eine Suche mit *kleiner als... (<)* und *größer als ...(>)* möglich, z. B. `>2500`. Das Eingabefeld des Filters färbt sich bei der Eingabe (mit der EINGABE-Taste) rot, wenn eine Filtereingabe nicht als numerischer Wert verwendbar ist.

Feed

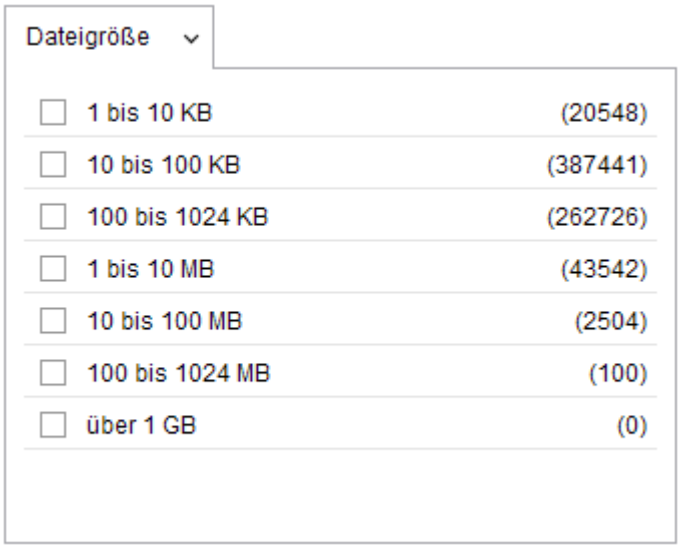
Der Filter für die Suche im Feed dient der Suche über alle Feed-Inhalte. Dies sind sowohl manuell erstellte Beiträge und Kommentare als auch automatisch generierte Einträge, wie z. B. zu einer neuen Dokumentversion. Im Feed sind damit auch die Versionsnummern und Versionskommentare enthalten, sowohl zu der aktuellen Arbeitsversion eines Dokuments als auch zu älteren Versionen.

In der Liste des Feed-Filters werden die am häufigsten verwendeten Hashtags als Suchvorschläge angezeigt. Diese werden wie alle anderen Eingaben auch in diesem Filter nach den üblichen linguistischen Regeln im Textinhalt des Feeds gesucht.



Dateigröße

Der Filter für die Dateigröße eines Dokuments ist der einzige ohne Eingabefeld. Die Auswahl des Größenbereichs erfolgt über eine Liste.

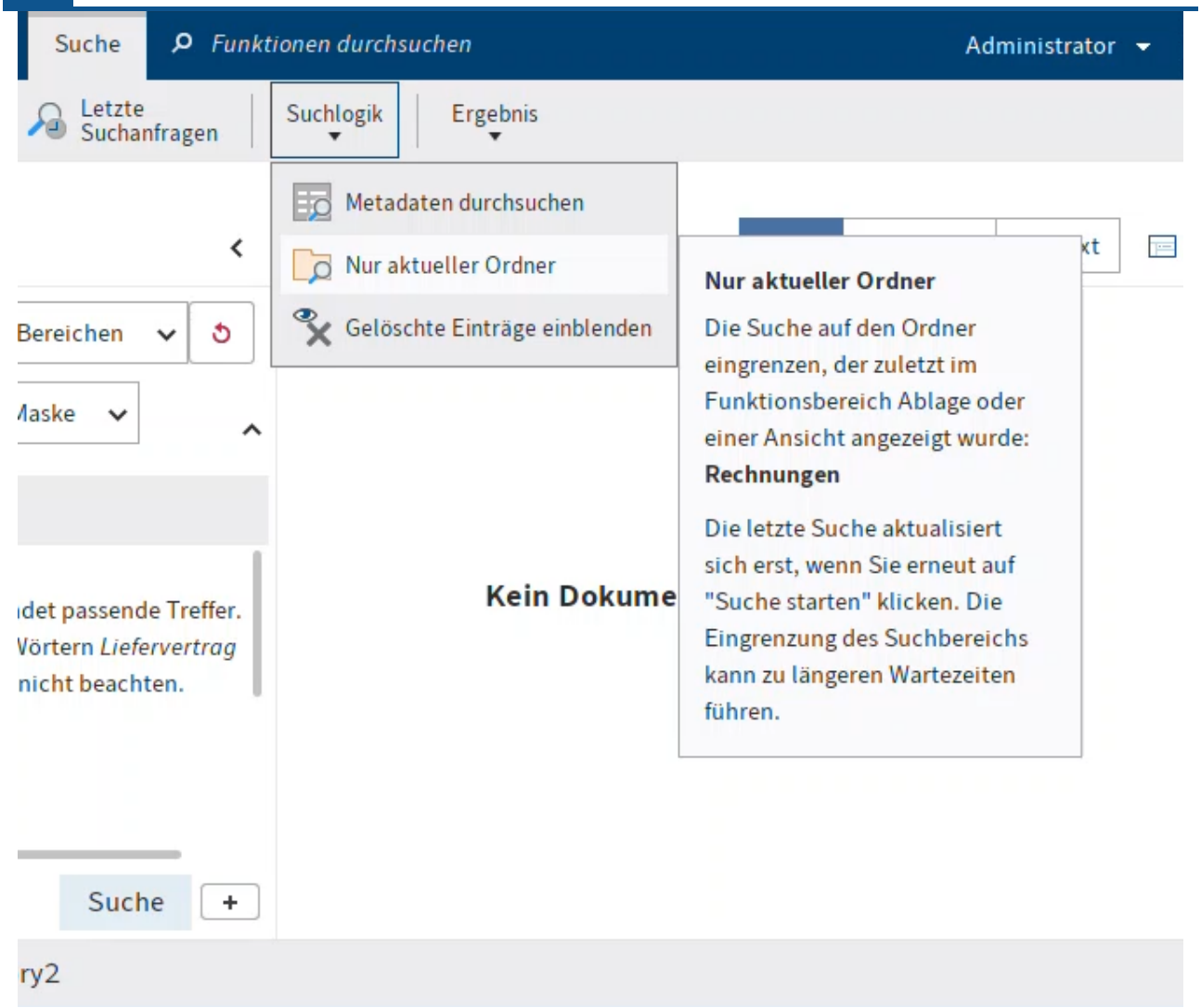


Space durchsuchen

Der Filter für ELO Spaces ermöglicht eine Suche in einem ausgewählten Space (Teamspace oder Workspace).

Suche im aktuellen Ordner

Die Suchoption *Nur aktueller Ordner* befindet sich im Menüband und grenzt die ELO iSearch auf den zuletzt benutzten Ordner ein – im Tooltip wird der aktuelle Ordner angezeigt.



Da es sich bei dieser Einstellung um eine nachträgliche Filterung der Treffer aus dem Suchindex handelt, verlangsamt dies die Suche deutlich. Es ist daher sinnvoll, diese Einstellung nur dann einzusetzen, wenn bereits eine relativ kleine Treffermenge über andere Suchkriterien sichergestellt ist. Die Suchfilter bieten sich dafür an.

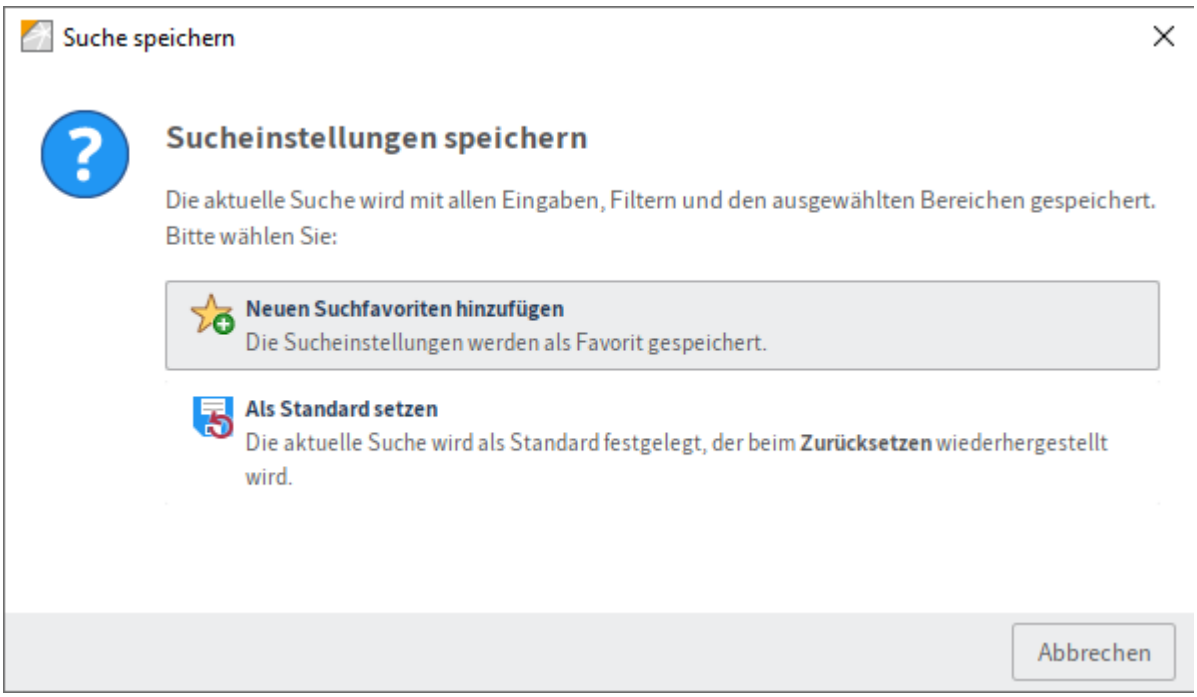
Suche speichern

In der ELO iSearch ist es möglich, Sucheingaben zu speichern. Verwenden Sie dazu die Funktion *Suche speichern* (*Menüband > Suche*).

Mit dem Speichern einer Suche halten Sie die gesamte ELO iSearch-Sucheingabe fest, inklusive der Auswahl der Filter und Filterwerte. Die Einstellung *Nur aktueller Ordner* (im Menüband) wird dabei nicht abgespeichert, da diese Auswirkungen auf die Such-Performance hat und daher nur in Ausnahmefällen verwendet werden sollte.

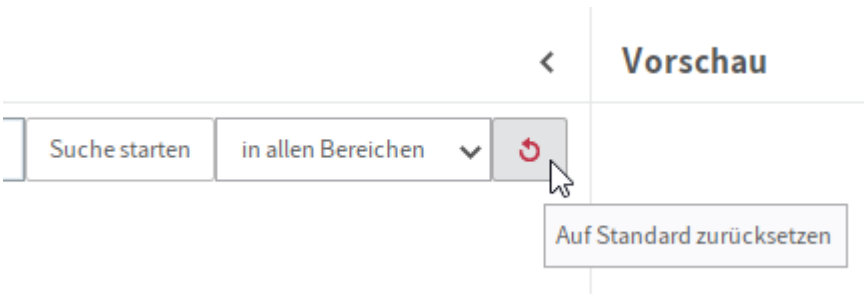
Sie können sich z. B. einen Suchfavoriten für die Recherche nach aktuellen E-Mails anlegen. Dann müssen Sie nur noch das jeweilige Suchwort eingeben und haben Maske und Zeitraum bereits richtig vorbelegt.

Beim Speichern der Suche werden zwei Möglichkeiten angeboten: das Anlegen eines neuen Suchfavoriten oder die Anpassung der Standardfilter.



Standardfilter

Die in Kapitel Filter beschriebenen Standardfilter können Sie per *Suche speichern* an Ihre die jeweiligen Vorlieben anpassen. Der somit für Sie definierte persönliche Standard wird beim Start des Clients verwendet und lässt sich jederzeit über die Button *Auf Standard zurücksetzen* wiederherstellen.

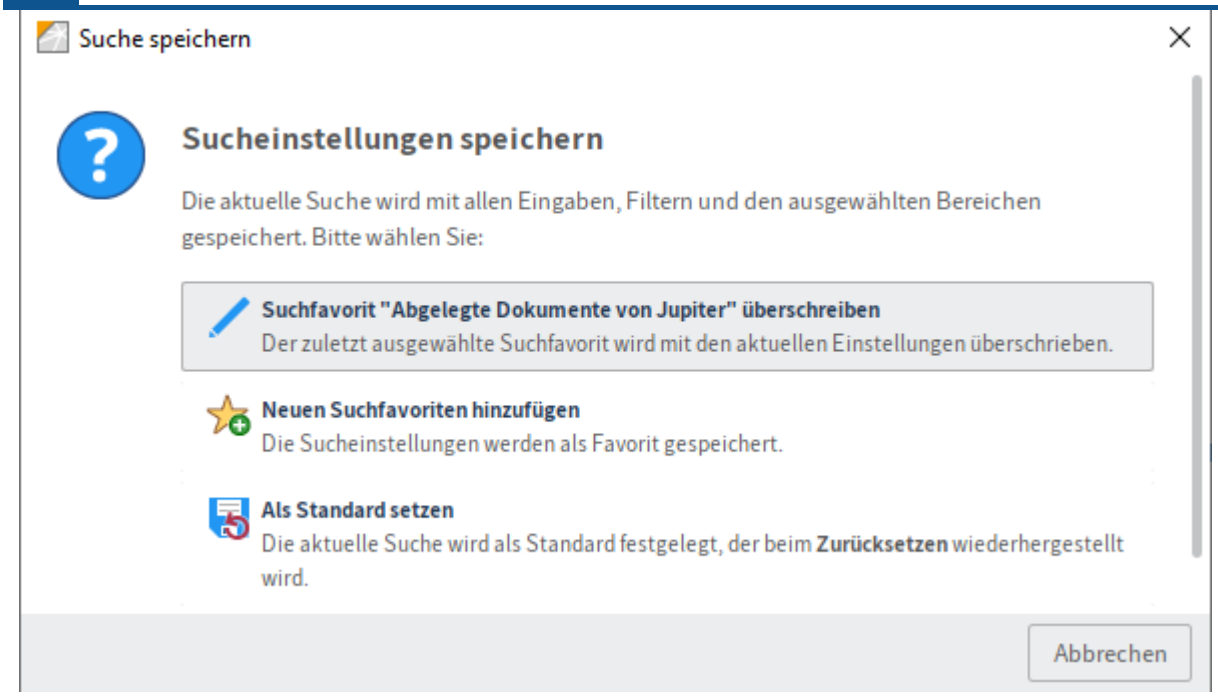


Wird gerade ein Suchfavorit verwendet und angepasst, wird zunächst auf den gespeicherten Zustand des Suchfilters zurückgestellt. In diesem Fall gelangen Sie mit einem zweiten Aufruf von *Auf Standard zurücksetzen* zurück den Standardfiltern.

Suchfavoriten

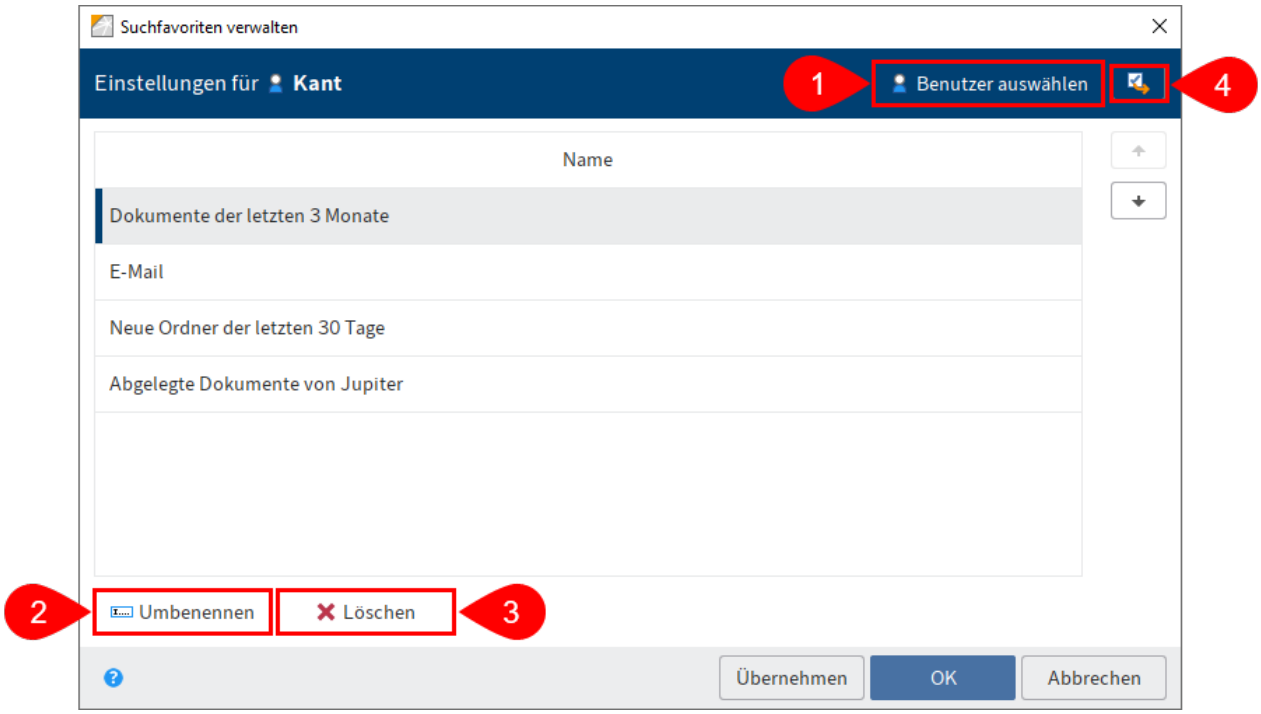
Suchfavoriten geben Ihnen die Möglichkeit, verschiedene wiederkehrende Rechercheaufgaben zu vereinfachen. Die Favoriten lassen sich im Menüband einblenden. Diese erscheinen nach einem Neustart dann auch als Kacheln im Bereich *Mein ELO*, clientübergreifend im ELO Java Client und ELO Web Client. Beim Anlegen eines Suchfavoriten werden Sie aufgefordert, einen Namen bzw. eine eindeutige Bezeichnung zu vergeben.

Haben sie gerade einen Suchfavoriten in Verwendung und speichern dann die Suche, so bietet Ihnen der Dialog zusätzlich die Möglichkeit, auch den aktuellen Suchfavoriten anzupassen:



Suchfilter administrieren

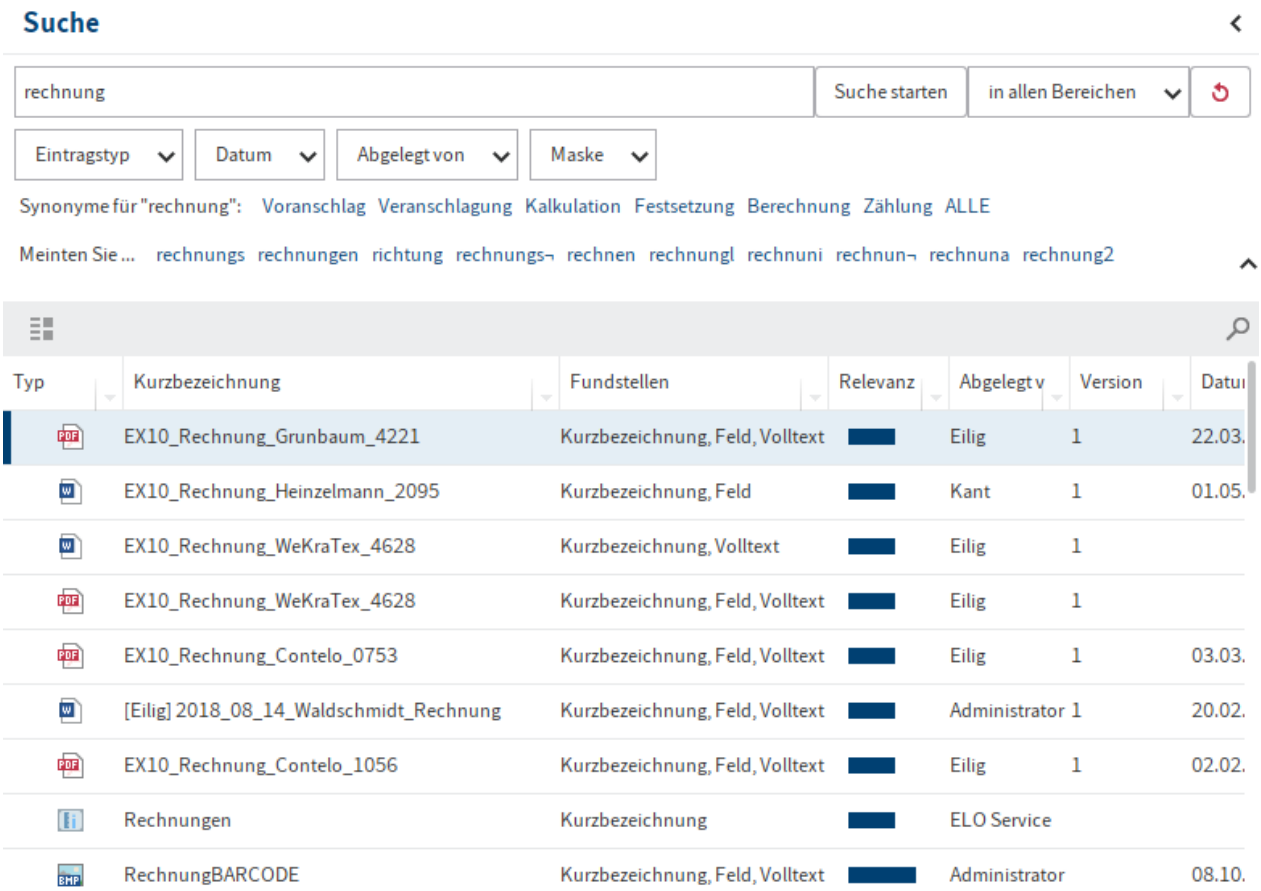
Die Administration kann globale Suchfilter für andere ELO Konten definieren. Navigieren Sie im Tab *Suchfavoriten* zur Funktion *Suchfavoriten verwalten*.



Über *Benutzer auswählen* (1) können Sie auswählen, wessen Favoriten verwaltet werden sollen. Sie können bei einem Konto oder bei einer Optionengruppe Suchfavoriten umbenennen (2) oder löschen (3). Mit *Einstellungen auf andere Benutzer übertragen* (4) können Sie im Dialog ausgewählte Suchfavoriten an andere Konten, Optionengruppen oder global verteilen. Hierfür wird auch die Einstellung der Standardfilter im Dialog eingeblendet, wenn diese beim ausgewählten Konto gespeichert wurde.

Trefferliste

Als Ergebnis der Suche wird eine Tabelle (Trefferliste) mit Einträgen angezeigt, die Ihrer der Suchanfrage entsprechen. Die Sortierung der Einträge erfolgt chronologisch absteigend nach ihrem Ablagedatum. Das bedeutet, dass die neuesten Dokumente in der Trefferliste oben stehen. Diese Sortierung der Trefferliste kann jederzeit angepasst werden, entsprechend der jeweiligen Bedürfnisse und der für das Suchergebnis relevanten Werte und Informationen.



Information

Bei der Anzeige der Fundstellen in der Tabellenansicht (siehe entsprechende Spalte), werden alle zur Sucheingabe passenden Fundstellen für das gefundene Dokument angezeigt, auch wenn eine konkrete Fundstelle im Suchbereich (siehe Abschnitt Suchbereich/Durchsuchte Felder) ausgeschlossen worden war. Das betreffende Dokument wurde jeweils nur über den eingestellten Suchbereich gefunden, jedoch kann es hilfreich sein, dass auf weitere im Dokument vorkommende Treffertexte aus allen Feldern hingewiesen wird.

Information

Fundstellen werden in der Trefferliste prinzipiell nur angezeigt, wenn über das Sucheingabefeld eine Textsuche durchgeführt wurde (ggf. auch in Kombination mit Filtern). Wenn dagegen das Sucheingabefeld leer ist oder das Wildcard-Zeichen * enthält, weil z. B. ausschließlich über Felder gefiltert wird, können keine Fundstellen angezeigt werden.

Treffermenge und -seiten

Um die Anzeige der Treffermenge weiter zu beschleunigen, gibt es einen Parameter, der die maximale Anzahl an Treffern eingrenzt, die angezeigt werden sollen. Dieser Parameter ist serverseitig konfigurierbar und kann den Anforderungen entsprechend angepasst werden. Hierfür ist die entsprechende Berechtigung notwendig.

Wird diese Anzahl bzw. dieser Parameterwert überschritten, werden lediglich ältere Dokumente nicht im Suchergebnis angezeigt. Durch Verwendung eines Datumsfilters können Sie gezielt auch auf diese Dokumente zugreifen oder aber ganz bewusst ausschließen.

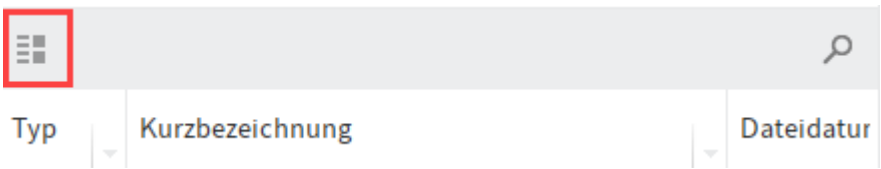
Der ELO Java Client zeigt die Treffer aufgeteilt in einzelne Trefferseiten an. Dies ist notwendig, da die Aufbereitung der Ergebnisse linear von der Anzahl der Treffer abhängt. Mit den Pfeilen können Sie seitenweise weiter- oder zurückblättern.

Beachten Sie

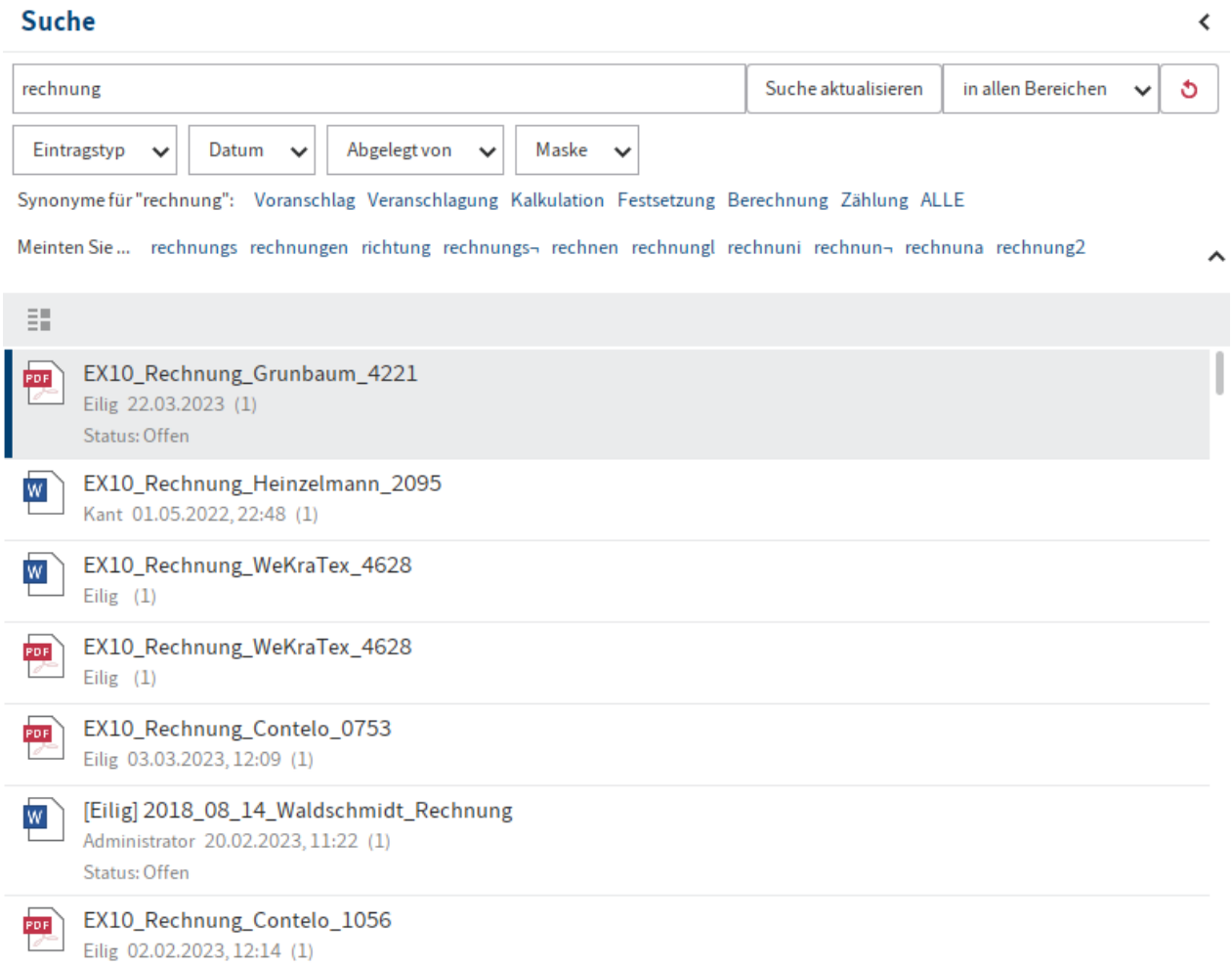
Die angezeigte Gesamttrefferanzahl ist nur eine Schätzung. Die Berechtigungen in ELO können sich durch fachliche Anwendungen zur Laufzeit ändern (z. B. das Ändern von Berechtigungen oder das Ablegen von neuen Dokumenten in ELO). Der Indexserver filtert daher alle Einträge nach, wodurch die angezeigte Treffermenge geringer oder größer ausfallen kann. Der Client passt den angezeigten Wert an, sobald das Trefferende erreicht und die Nachfilterung abgeschlossen ist. Auch die Suche im aktuellen Ordner verwendet eine nachträgliche Filterung. Hier weichen die Werte deutlich ab, da sich die Schätzung immer auf das ganze Repository bezieht.

Darstellung als Liste, Kachel oder Baum

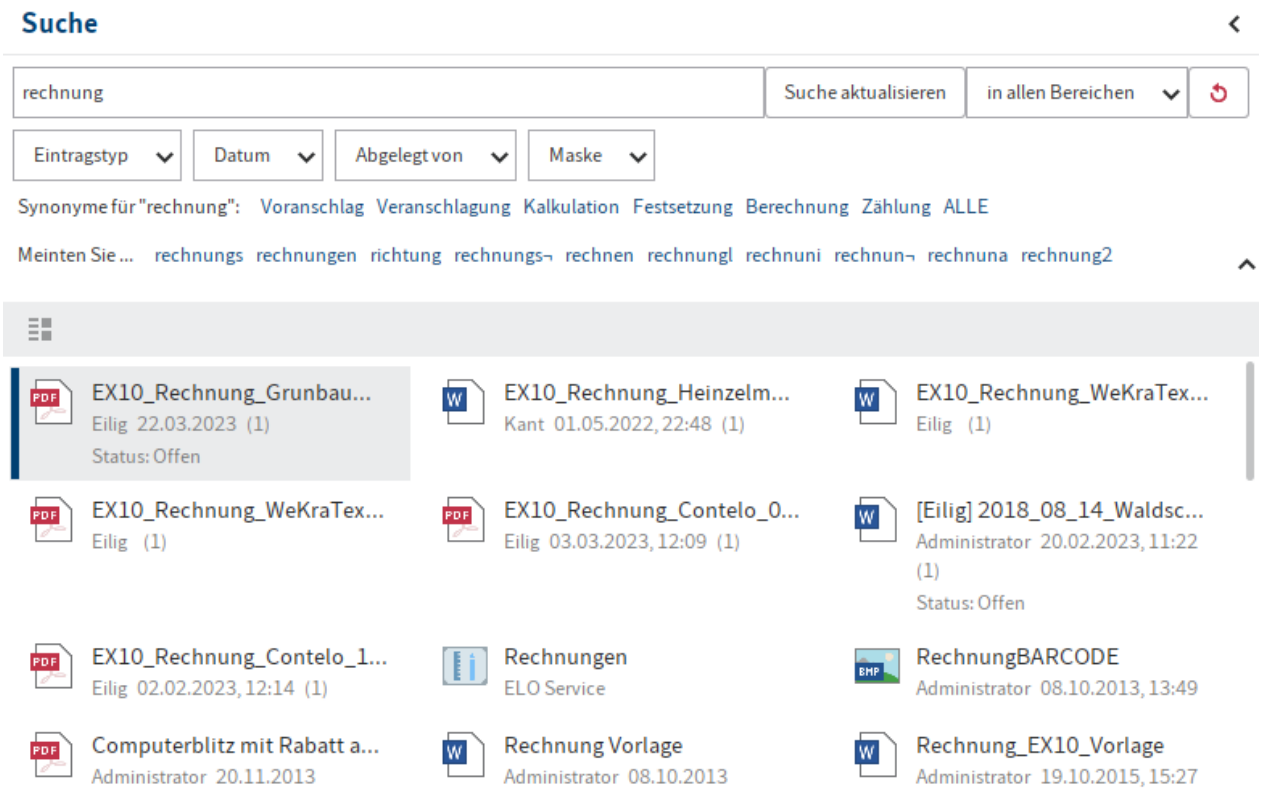
Der ELO Java Client bietet bei der Darstellung der Suchtreffer neben der Tabellenansicht auch eine Listendarstellung, eine Kachel- oder eine Baumansicht an. Die Darstellungsoptionen finden Sie in der grauen Funktionsleiste.



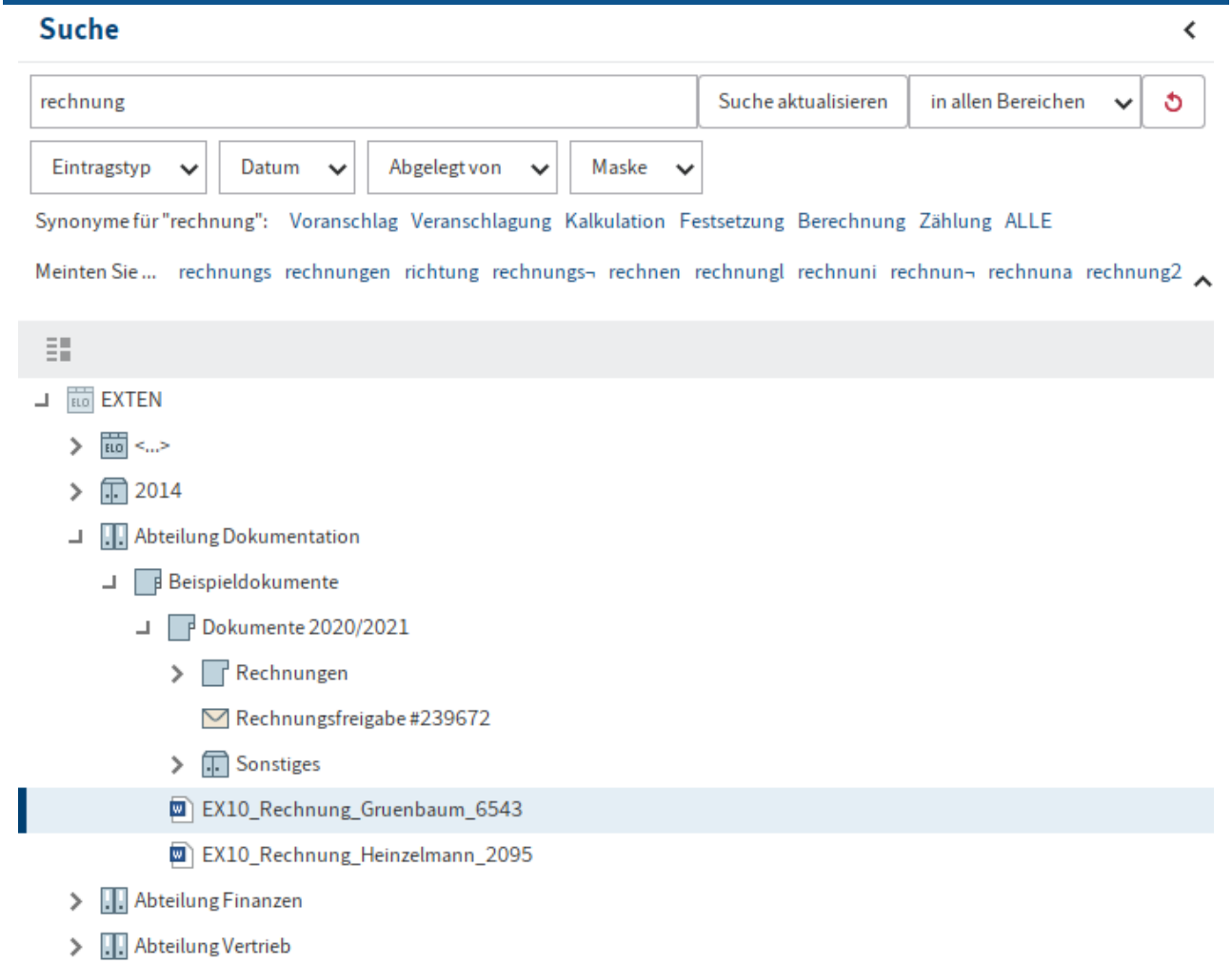
Die Listendarstellung ist eine besonders kompakte Darstellung der Suchergebnisse, die sich gut für eine übersichtliche Anzahl von Treffern eignet und dadurch viel Platz für die Vorschau der Dokumente auf dem Bildschirm lässt.



Die Kacheldarstellung eignet sich vor allem für die Anzeige von Miniaturansichten. Zu beachten ist hierbei, dass diese vom Client erzeugt werden, dafür also alle Dokumente zunächst vom Server geladen und bearbeitet werden müssen. Diese Darstellung eignet sich für überschaubare Treffermengen im lokalen Netzwerk.

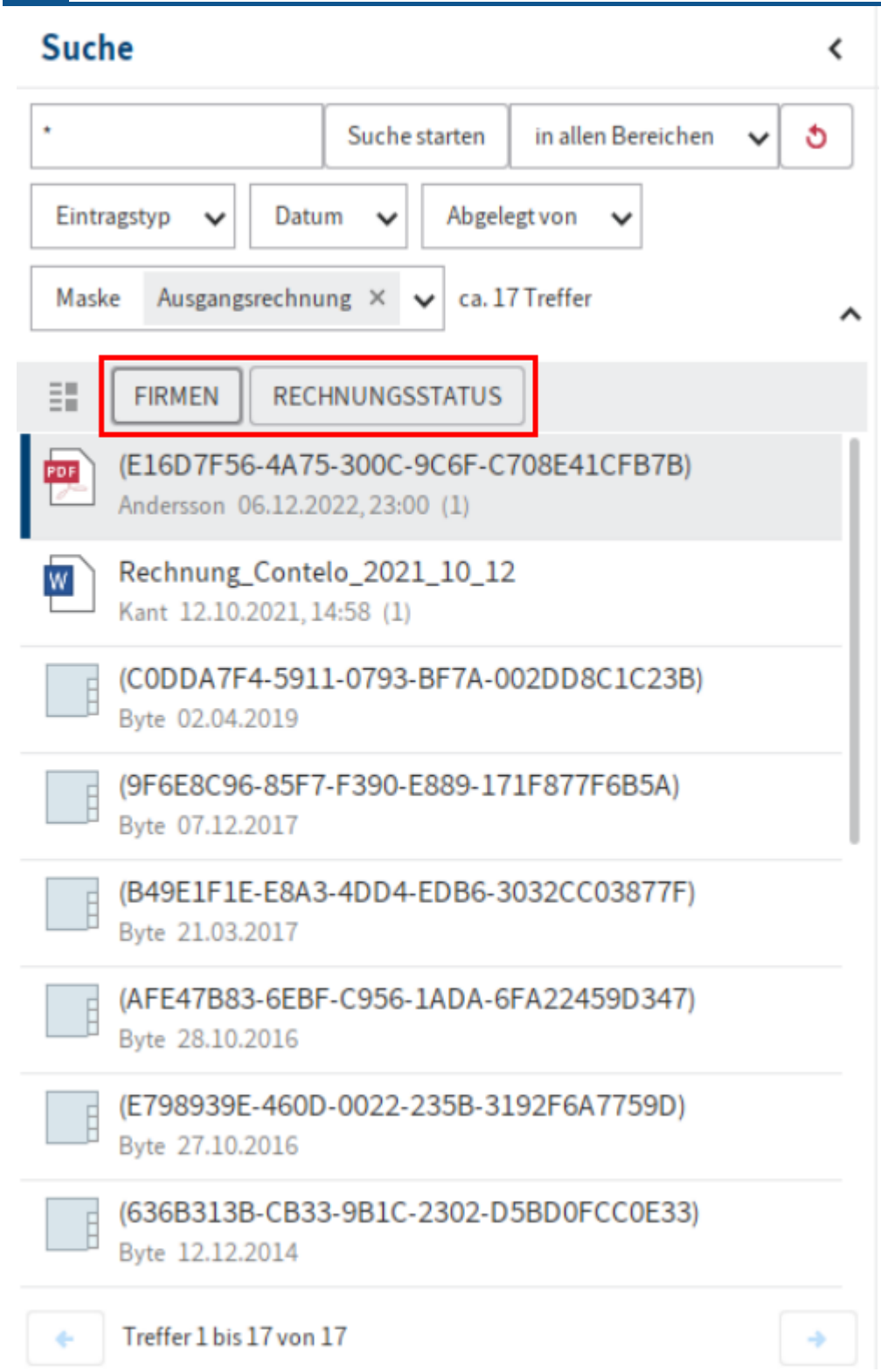


Die Baumdarstellung ermöglicht es, die Treffer nach auswählbaren Kriterien in einer Baumstruktur zu ordnen. Eine vordefinierte Möglichkeit ist die Darstellung entsprechend der Struktur in ELO. Es können auch eigene Strukturen anhand der Felder der Einträge definiert werden. Für eine Maske *Rechnung* wäre z. B. eine Struktur nach Jahr und Firmenname möglich, wenn die benötigten Werte in der Maske als Felder vorhanden sind.



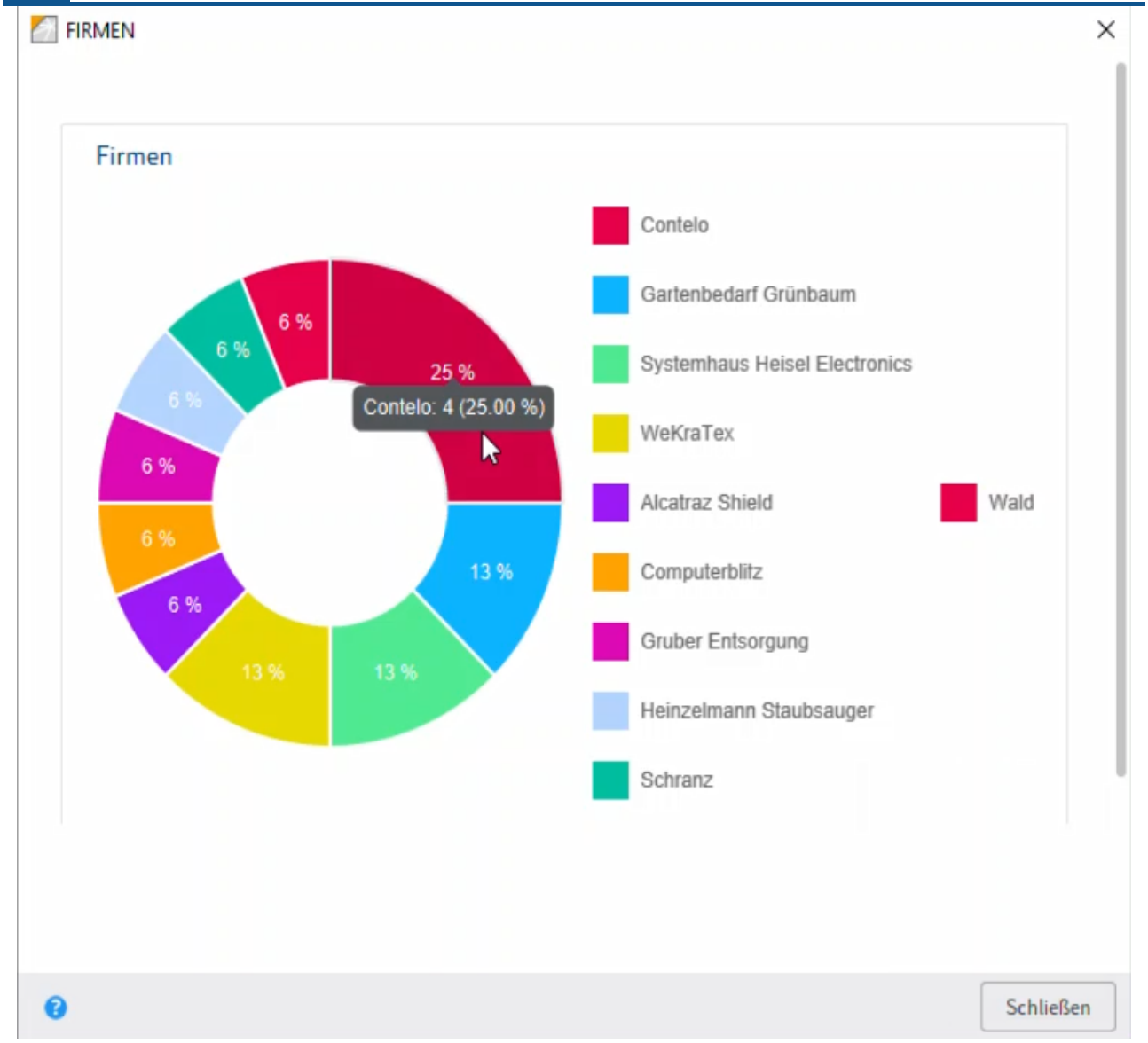
Dashboards

Im Funktionsbereich *Suche* stehen Ihnen nach einer Suche mit einem Maskenfilter im Suchergebnis Buttons zur Verfügung, über die Sie Dashboards aufrufen können. Diese Buttons werden nur angezeigt, wenn in der jeweiligen Maske Dashboards hinterlegt wurden.



Sie können die Dashboards in der ELO Administration Console hinterlegen.

Durch die Dashboards werden Metadaten der Einträge, die unter einer Maske abgelegt wurden, unterschiedlich visualisiert. Es können Balken-, Linien- und Doughnut-Diagramme angezeigt werden.



Mit der Auswahl eines Dashboard-Buttons erscheint ein neues Dialogfenster. Per Mouseover können Sie sich Details zu den Dashboards-Diagrammen anzeigen lassen.

Informationen zur Erstellung von Dashboards finden Sie in der Dokumentation *ELO Pakete* unter *Metadaten > Masken > Masken konfigurieren > Dashboards erstellen*.

Technische Beschreibung

Architektur

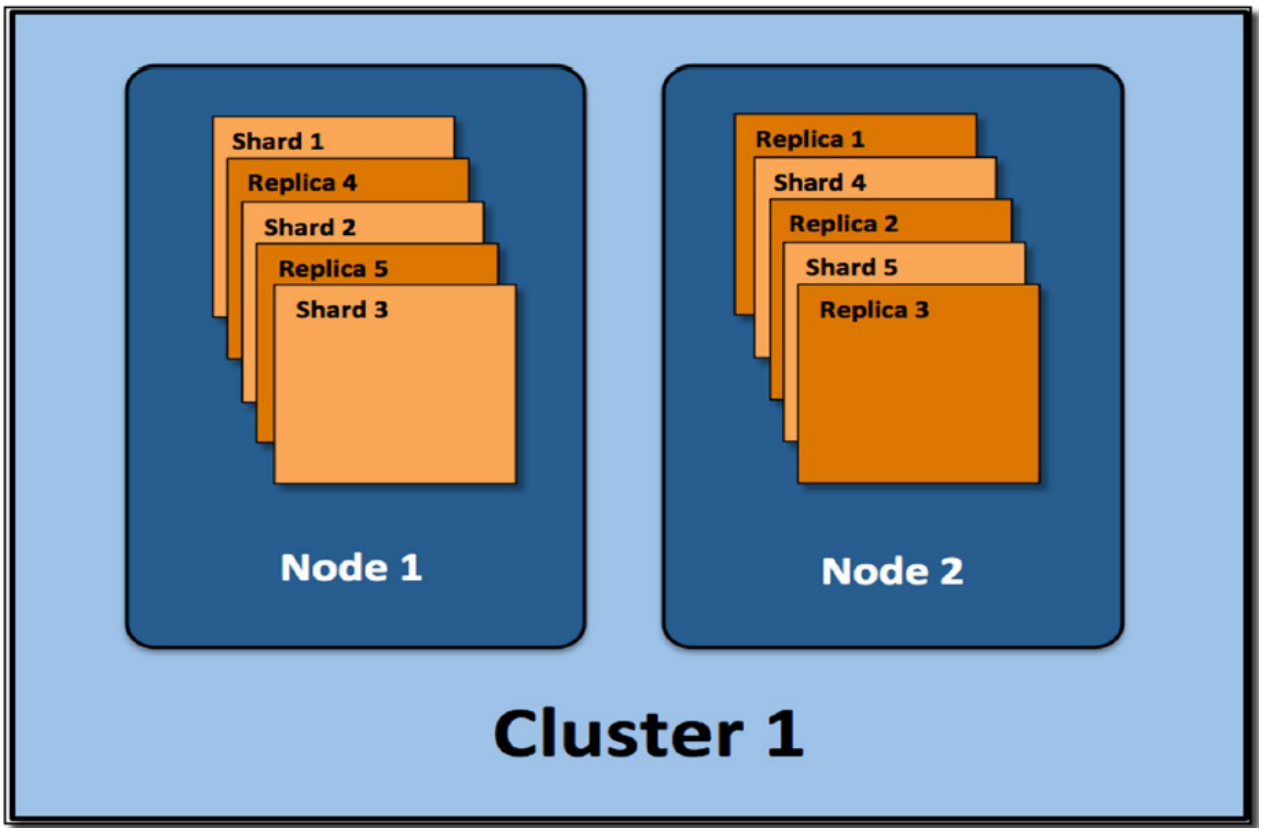
Elasticsearch

Die Elasticsearch ist eine verteilte, hoch skalierbare Open-Source-Such-Engine auf Basis von Apache Lucene. Mit der Elasticsearch ist es möglich, große Datenmengen zu speichern und extrem schnell zu durchsuchen. Die Kommunikation erfolgt dabei über ein RESTful Webinterface.

Durch die große Verbreitung können Sie im Internet viele Grundlagendokumente und Vorgehensweisen finden. In dieser Dokumentation werden deshalb lediglich die Details herausgearbeitet, die relevant für den Einsatz im ELO Systemumfeld sind. Für Informationen darüber hinaus empfehlen wir folgenden Einstiegspunkt für weitere Recherchen: https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html.

Lucene

Die Elasticsearch nutzt Apache Lucene als *core library*. Mit der Elasticsearch kann jeder Index in mehrere Stücke (sogenannte *Shards*) aufgeteilt werden. Ein Shard entspricht einem Lucene-Index. Dieser besteht physikalisch aus einem Ordner mit den zugehörigen Index-Dateien.



Die Shards können auf mehrere Server (*Nodes*) aufgeteilt werden. Es können *Replicas* gebildet werden. Dies ergibt dann einen *Cluster*.

Sprachunterstützung

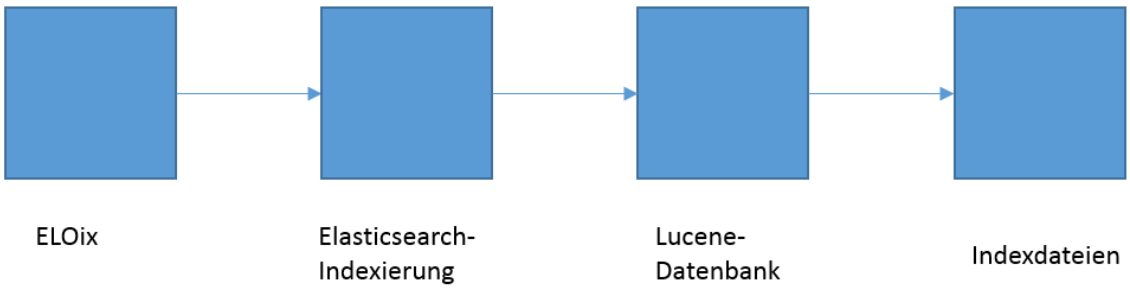
Die Elasticsearch verwendet sowohl bei der Indexierung als auch bei der Suche eigene Sprach-Analyzer. Dadurch werden sprachspezifische Suchregeln und -muster optimiert. Bei den ELO Spezifika müssen Sie daher beachten, dass die Repository-Sprache (Auswahl bei der Installation) einen maßgeblichen Einfluss auf die Suche besitzt.

Folgende Sprachen werden unterstützt:

Deutsch, Englisch, Französisch, Spanisch, Italienisch, Portugiesisch, Dänisch, Schwedisch, Polnisch, Niederländisch, Tschechisch, Ungarisch, Rumänisch, Türkisch, Bulgarisch, Finnisch, Griechisch, Norwegisch, Russisch.

Verarbeitung mit ELO

Die Verarbeitung mit ELO können Sie dem folgenden vereinfachten Schema entnehmen:



Der ELO Indexserver übergibt die Daten an die Elasticsearch, die bei der Indexierung (und auch bei der Suche) eigene Sprach-Analyzer verwendet. Die Indexdaten werden in Lucene-Indexen gespeichert.

Für das Suchen im Index wird also der Zugriff auf:

- die Index-Dateien,
- eine konsistente Lucene-Datenbank,
- ein funktionsfähiger Elasticsearch-Sprach-Analyzer,
- ein laufender ELO Indexserver (ELOix)

vorausgesetzt.

Sicherheit

Seit der ELO ECM Suite 10.2 erfolgt die Kommunikation zwischen dem ELO Indexserver (ELOix) und der Elasticsearch verschlüsselt (SSL). Dazu wird das *Search Guard Plugin* verwendet.

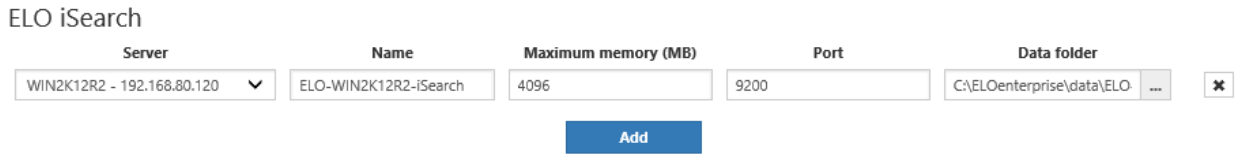
Installation und Upgrade

ELO Server Setup

Die Installation oder ein Upgrade erfolgt über das ELO Server Setup. Dieses erzeugt – falls noch nicht vorhanden – die Konfigurationen, installiert das Elasticsearch-Programm und den Dienst,

sowie ein Datenverzeichnis, in dem die Indexdaten abgelegt werden. Folgende Optionen können im ELO Server Setup gesetzt werden:

- Name des Dienstes
- Memory-Wert
- Port
- Datenverzeichnis



- Mit ELOenterprise auch weitere Elasticsearch-Server (siehe dazu auch Lizenzbestimmungen bei der Verteilung von Serverprozessen).

Beachten Sie

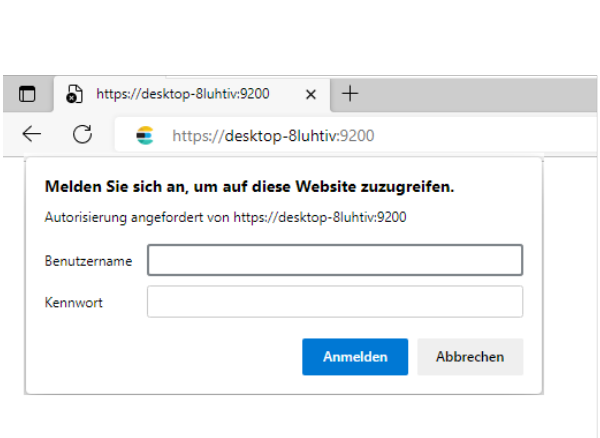
Ab ELO 21.2 wird der Port der ELO iSearch (Standard: 9200) mit TLS/SSL gesichert.

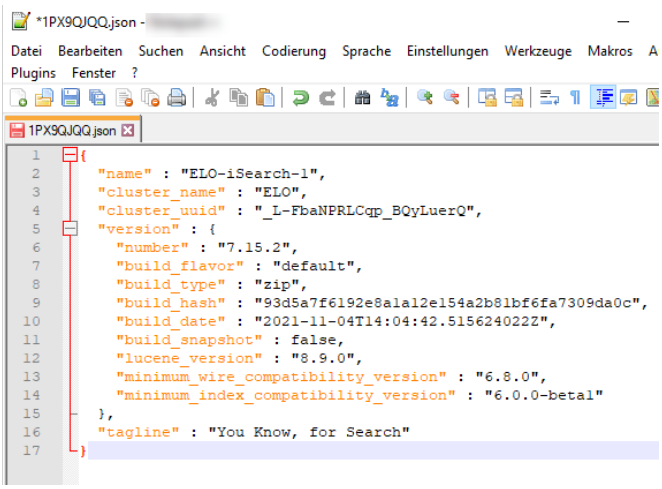
Bei Aufruf über einen Browser wird beim ersten Zugriff ggf. eine Zertifikatswarnung angezeigt.

Je nach verwendetem Browser und Betriebssystem müssen Sie das angebotene Zertifikat entweder in den Zertifikatsspeicher des Browsers oder in den Zertifikatsspeicher des Betriebssystems übernehmen. Beispielsweise laden Sie bei Verwendung von Microsoft Edge oder Google Chrome unter Microsoft Windows die Datei *keystore.jks* aus `<ELO>\config\elastic\<Server ELO iSearch>\certificates` in den Windows-Zertifikatsspeicher. Bei Verwendung von Mozilla Firefox nehmen Sie das Zertifikat durch Bestätigen der Warnmeldung in den Zertifikatsspeicher von Firefox auf.

Prüfungen nach der Installation

Nach der Installation können Sie zur Überprüfung mit administrativem Tomcat-Konto `https://<server>:<port>` (Standard-Port ist 9200) aufrufen. Bei Verwendung eines beliebigen Browsers wird eine JSON-Datei ausgegeben, die Sie in einem Editor öffnen können.

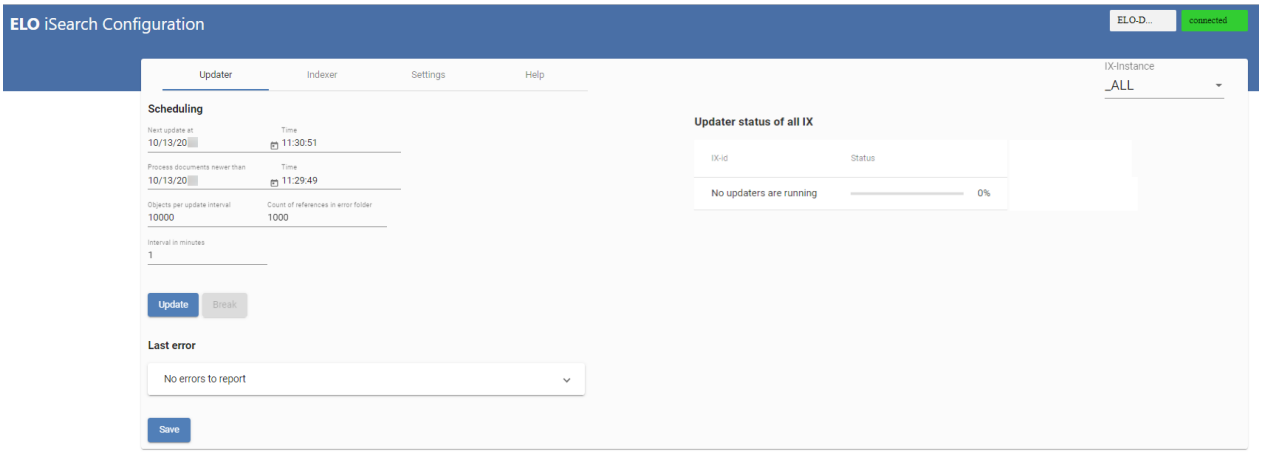




Dadurch können Sie sich wichtige Basisdaten wie Version und Namen anzeigen lassen.

Über die ELO iSearch Configuration können Sie weitere Prüfungen vornehmen. Sie erreichen die Konfiguration über die Indexserver-Konfigurationsseite oder über folgenden Link (Anmeldung mit administrativem Tomcat-Konto):

`https://<server>:<ixport>/ix-<REPOSITORY>/manager/esconfig/#/iSearchConfig`

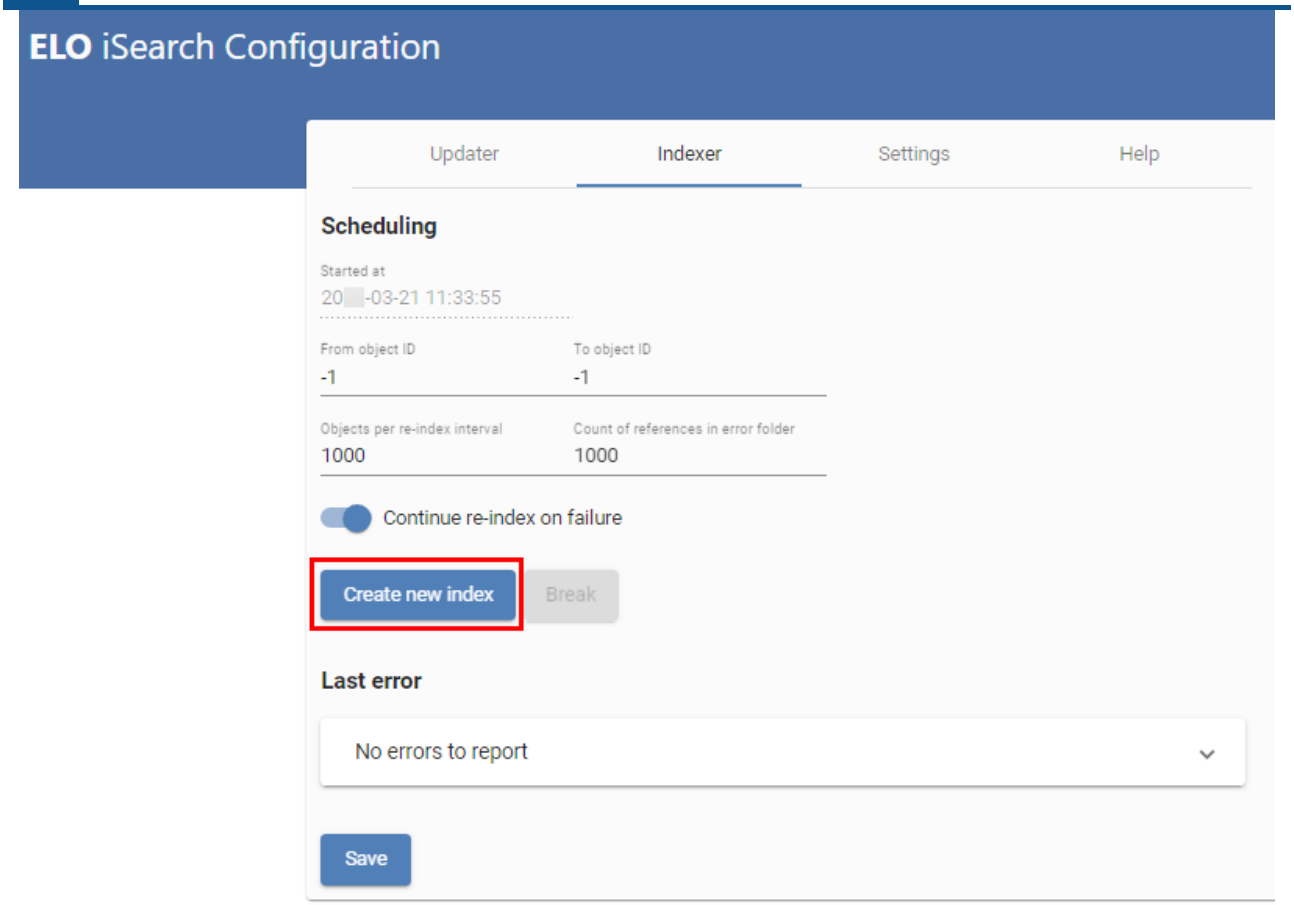


Indexaufbau nach Upgrade von älteren ELO Versionen

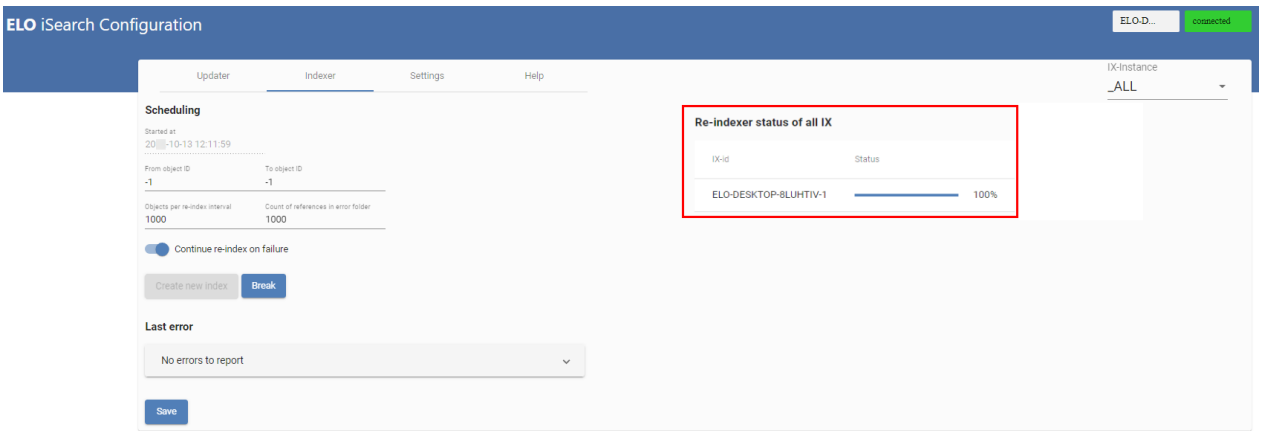
Nach einem Upgrade einer früheren ELO Version (bis einschließlich 21.01) auf Version 23 LTS ist ein Neuaufbau (Re-Index) der Indexdaten erforderlich. Das ELO Server Setup entfernt die bisherigen Elasticsearch-Indexe, da diese mit der aktuellen Elasticsearch-Version nicht mehr kompatibel sind.

Bei der Re-Indexierung wird zunächst der vorhandene Index gelöscht und dann neu aufgebaut. Alle Felder werden sowohl *tokenized* (in Token aufgeteilt) als auch *not tokenized* (als Phrase) abgespeichert, sodass eine Änderung an dieser Feld-Einstellung keinen Re-Index erfordert. Zusätzlich werden aus den durch den ELO Volltext-Prozess entstandenen *FT*.txt*-Dateien die Volltextinhalte ausgelesen und für die Indexierung verwendet.

Ein Re-Index wird über die ELO iSearch-Konfiguration im Tab *Indexer* angestoßen.

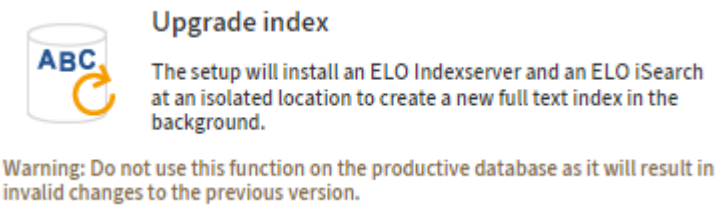


Der Fortschritt wird nach einem Refresh der Seite angezeigt:



Indexaufbau im Hintergrund

Bei Upgrades ab der ELO Version 9.3 ist es möglich, über das ELO Server Setup mit der Funktion *Upgrade Index* einen isolierten ELO Indexserver zu installieren, der den Elasticsearch-Index im Hintergrund aufbaut. Dies ist vor allem bei großen Systemen sinnvoll, bei denen die Re-Indexierung mehrere Tage dauert. Diese Funktion können Sie sowohl bei einem Update zu einer aktuelleren ELO Version verwenden als auch bei Änderungen in den Einstellungen der iSearch.



Diesen Index können Sie dann bei einem ELO Upgrade übernehmen, sodass direkt nach dem Upgrade die volle Suchfunktionalität zur Verfügung steht.

Weitere Details zum Verfahren finden Sie im Abschnitt Indexaufbau im Hintergrund.

Sizing und Performance

Eine performante Suche stellt die Grundlage zur Akzeptanz des ELO Systems auf Benutzungsebene dar. Daher müssen Sie neben der Optimierung der Datenbank-Performance auch regelmäßig (z. B. halbjährlich, in Abhängigkeit von der hinzukommenden Dokument- und Datenmenge) eine Optimierung der Elasticsearch durchführen. Die Notwendigkeit sollte zum einen auf Benutzungsebene überprüft werden (durch Ausführung komplexerer Suchen), zum anderen
aber auch auf Serverebene (durch gelegentliche Log-Datei-Analysen, siehe Abschnitt zu Log-Datei).

Allgemeine Aussagen zum Sizing

Die Performance der Elasticsearch wird durch folgende Faktoren beeinflusst:

- Arbeitsspeicher, sowohl für die JVM (Java Virtual Machine) der Elasticsearch, weiteren Speicher der Elasticsearch für Caches, der vom Betriebssystem verwaltet wird, als auch für das Betriebssystem allgemein
- Anzahl der Prozessoren/Kerne
- Speichertechnologie bzw. Speicher-Zugriffszeiten

Folgende generelle Aussagen zur Dimensionierung können gemacht werden:

- Arbeitsspeicher (siehe auch den Abschnitt Dimensionierung Arbeitsspeicher):
  - Für kleine Umgebungen sind die als Standard definierten 4 GB + 4 GB weiterer freier RAM ausreichend.
  - Für mittlere Umgebungen (ca. 10 Mio. Dokumente) sollten ca. 24 GB (12 GB + 12 GB) eingeplant werden.
  - Für große Umgebungen gibt es eine Obergrenze von rund 30 GB je JVM, hier ist auf jeden Fall eine Verteilung auf mehrere Server (Cluster) sinnvoll.
- Prozessoren/Kerne:
  - Für jede Suche wird ein Thread benötigt, das heißt durch eine höhere Anzahl an Prozessoren/Kernen wird eine bessere Parallelisierung erreicht.
- Festplatten:
  - Schnelle Festplatten (SSD) machen sich signifikant bemerkbar. Storage auf SSD-Basis ist für den Betrieb einer Elasticsearch dringend empfohlen.

Auch bereits für mittlere Umgebungen empfehlen wir den Einsatz von ELOenterprise (siehe Lizenzbestimmungen bei der Verteilung von Serverprozessen).

Dimensionierung Arbeitsspeicher

Bei der Elasticsearch hängt der Speicherverbrauch nicht nur mit der Anzahl und Größe der Dokumente zusammen, sondern zu einem wesentlichen Teil mit der Art der Nutzung durch die Anwender.

Generell muss man beim Speicherverbrauch zwischen dem Java-Heap für den ELO iSearch-Dienst und dem darüber hinaus benötigtem RAM für die Elasticsearch unterscheiden. Hier gibt es die klare Empfehlung, mindestens so viel freien Speicher zusätzlich bereitzuhalten, wie für den Java-Heap konfiguriert ist.

Den Java-Heap konfigurieren Sie fest beim ELO iSearch-Dienst, beispielsweise 10 GB, bei einer neuen ELO Installation initial 4 GB. Dann sollte ebenso viel freier Speicher, der nicht von anderen Applikationen und dem Betriebssystem selbst belegt ist, dem Betriebssystem als Filesystem-Cache für die Elasticsearch zur Verfügung stehen. Die Elasticsearch nutzt diesen Speicher, in dem er vom Betriebssystem als Filesystem-Cache benutzt wird. Inhaltlich sind dort hauptsächlich Indexdaten (fix und temporär) und manche Arten von Caches für Suchanfragen zu finden.

Im Java-Heap befindet sich dagegen die Elastic-Applikation selbst, dazu unter anderem eine Grundmenge Speicher je laufendem Shard (Lucene-Index), sowie der Query-Cache und der Fielddata-Cache.

Für die initiale Größe des Java-Heaps könnte man folgenden Richtwert angeben, der später je nach Nutzungsverhalten angepasst werden muss: 4 GB Grundmenge und dann zusätzlich noch so viel, wie gerade benötigt wird (1 GB für 1 Mio. ELO Dokumente ist ein guter Anfangswert). Alles Weitere ist als Filesystem-Cache viel wertvoller. Als Erfahrungswert wird bei ca. 10 Mio. Dokumenten die richtige Einstellung des Java-Heaps bei 8 bis 16 GB liegen. Der genaue Wert hängt vom Nutzungsverhalten und der Art der Indexfelder ab. Wenn z. B. die Anwender im ELO Client häufig Filter für Indexfelder benutzen, die als *tokenized* markiert sind und sich sogenannte Kontextterme (d. h. Filter-Werte, die auswählbar sind für dieses Feld) anzeigen lassen und der Index für diese Maske Millionen Dokumente enthält, dann werden alle potenziell vorhandenen Feldwerte in den Java-Heap geladen und landen im Fielddata-Cache. Dies lässt die Größe des Cache und damit des benötigten Java-Heaps stark anwachsen. Dies passiert auch bei einer Sortierung auf solchen Feldern. Wenn dagegen entweder nur auf *notTokenized*-Feldern aggregiert wird oder keine Filtervorschläge durch Benutzer angefordert werden, beeinflusst dies den Fielddata-Cache nicht, sondern die Feldinhalte landen in temporären virtuellen Files, die dann im oben angesprochenen Filesystem-Cache des Betriebssystems zu finden sein sollten. Dies macht die Vorhersage schwierig, wie viel Java-Heap eingestellt werden sollte.

> **Beachten Sie**
>
> Für jede Maske wird ein separater Elasticsearch-Index aufgebaut. Dieser benötigt (für ein 1-Knoten-System ohne Replica) eine Grundmenge von etwa 50 MB Java-Heap. Eine große Anzahl von Masken erfordert daher einen großen Bedarf an Java-Heap und damit Arbeitsspeicher für die ELO iSearch.
>
> Die Empfehlung ist daher, dass bei der Konfiguration von Masken die Anzahl der Masken möglichst einen zweistelligen Wert nicht überschreiten sollte. Dies gilt sowohl für Masken der ersten als auch für Masken der zweiten Generation.

Für die gesamte Dimensionierung von Java-Heap und restlichem RAM als Filesystem-Cache empfehlen wir, bei einem System zunächst den Speicherverbrauch aller anderen Applikationen zu ermitteln, die noch auf dem gleichen Rechner oder der VM laufen, inklusive des Betriebssystems selbst. Beispielsweise läuft dort noch ein ELO Indexserver, ein ELO Textreader und ein ELO OCR, deren Verbrauch nicht allein durch den konfigurierten Speicher bestimmt ist, sondern ebenfalls eine große Menge an Filesystem-Cache erfordert. In einer Task-Manager-Applikation Ihres Betriebssystems lässt sich die Gesamtmenge für jede Applikation erkennen. Dann addieren Sie hierzu noch den Verbrauch des Betriebssystems. Der Gesamt-RAM des Systems setzt sich also zusammen aus diesem ermittelten Wert, dem konfigurierten Java-Heap für die Elasticsearch und mindestens gleich vielem freien Speicher für den Elastic Filesystem-Cache.

Beispiel:

ELOix/ELOtr/ELOocr/Betriebssystem benötigen z. B. 12 GB. Die ELO iSearch ist auf 8 GB Java-Heap konfiguriert und benötigt dazu noch 8 GB Filesystem-Cache. Die Gesamtbelegung ist also 12 GB + 8 GB + 8 GB = 28 GB. Damit hätte man dann bei einem 32 GB-System noch Platz für 4 GB. Das ist nicht besonders viel, wenn das Repository wächst.

Bei größeren Systemen empfiehlt es sich, die ELO iSearch auf einer separaten VM oder einem separaten Rechner zu betreiben, damit es nicht zu Konkurrenzsituationen um den Filesystem-Cache mit den anderen Applikationen kommt und das Betriebssystem zu Auslagerungsvorgängen gezwungen ist.

Generell gilt: Nur so viel Java-Heap wie nötig, aber so viel Filesystem-Cache wie möglich.

> **Information**
>
> Wir empfehlen grundsätzlich, den Java-Heap der ELO iSearch im ELO Server Setup einzustellen.
>
> Es gibt zusätzlich die Möglichkeit, den Java-Heap in der Datei *<ELO>\config\elastic\ELO-<Instanzname>\jvm.options.d\ELO-<Instanzname>.options* im ELO Verzeichnis einzustellen. Nach der Konfigurationsänderung müssen Sie die ELO iSearch neu starten. Die Änderungen gehen allerdings beim nächsten Durchlauf des ELO Server Setups verloren, deswegen ist diese Änderungsmöglichkeit nur für Testzwecke geeignet.
>
> Informationen zum Java-Monitoring der ELO iSearch zu Analyse- und Support-Zwecken finden Sie in der Dokumentation ELO Server > Wartung und Monitoring > Monitoring der Java-Umgebung > Server > ELO iSearch.

Optimierung der Indexierung

Sie können den Bereich der Indexierung nur bedingt optimieren. Die Anzahl der Dokumente zur gleichzeitigen Indexierung wird vom ELO Indexserver (ELOix) bestimmt. Die Erfahrungswerte für den Indexaufbau liegen im Bereich von 2-3 Millionen Dokumenten pro Tag inklusive Volltext-Dokumenten bei einer gut ausgestatteten Umgebung.

Wenn für ein ELO Dokument die Option In Volltext aufnehmen nicht aktiviert ist, wird kein Volltextdokument erzeugt und steht dementsprechend nicht zur Indexierung an. Der Re-Index lässt sich also beschleunigen, wenn weniger Volltextdokumente existieren. Auch die Größe des

Volltexts spielt für die Geschwindigkeit eine Rolle. Über die Option *maxFulltextContentMB* in den Indexserver Configure Options des ELO Indexservers können Sie die maximale Größe des zur Indexierung verwendeten Volltexts eines Dokuments bestimmen. Dokumente mit reinen Metadaten ohne Volltext sind bei der Re-Indexierung um ein Vielfaches schneller.

Optimierung der Suchgeschwindigkeit

Die Faktoren der Performance wurden schon in Allgemeine Aussagen zum Sizing beschrieben. In großen Umgebungen ergeben sich weitere positive Einflüsse auf die Suchgeschwindigkeit durch eine Erhöhung der Anzahl der Shards je Index, wenn diese auf mehrere Nodes verteilt werden. Sie müssen dabei beachten, dass eine Anpassung der Shard-Anzahl einen Indexaufbau (Re-Index) erfordert. Zur Konfiguration der Anzahl der Shards lesen Sie das Kapitel Festlegung der Shard-Anzahl pro Index.

Festplattenplatzbedarf für Indexe der Elasticsearch

Der Festplattenplatzbedarf für die Indexe der Elasticsearch hängt von verschiedenen Faktoren ab, die in jedem Kunden-Repository unterschiedlich stark vorhanden sind. Die Hauptfaktoren sind:

- die Anzahl der Dokumente und Ordner
- die Anzahl und Typen der kundenspezifischen und lösungsspezifischen Felder in Masken (Indexfelder)
- für jedes Feld die Menge der Daten, die jeweils enthalten sind. Hier spielt auch die Kardinalität des Felds eine Rolle. Bei Metadaten der Gen. 1 die Mehrwertigkeit über den sogenannten Spaltenindex, bei Metadaten der Gen. 2 die Eigenschaft, ob die jeweilige Aspektzuordnung die Kardinalität *MANY* hat (d. h. Kardinalität ist gleich *MANDATORY_MANY* oder *OPTIONAL_MANY*).
- die Einstellung, ob Volltext erzeugt und indiziert wird, der Anteil von Volltextdokumenten an der Gesamtanzahl der Dokumente und die Einstellung zur Größenbeschränkung des zu indizierenden Volltextes (*maxFulltextContentMB*)
- ab ELO Version 25.1 die Einstellung, ob AI-Vektoren für den Volltext erzeugt und indiziert werden sowie die Größe des Volltextes
- bei Metadatenfeldern vom Typ Relation werden (konfigurierbar) markierte weitere Felder des referenzierten Dokuments in das referenzierende Dokument mit aufgenommen
- bei Dokumenten, die in einer Region liegen (z. B. bei Dokumenten unterhalb eines Business-Objekts), werden (konfigurierbar) markierte weitere Felder des Regionenobjekts mit in die Dokumente aufgenommen, die in der Region liegen

Da diese Faktoren in jedem Kunden-Repository sehr individuell vorliegen, lässt sich für den zu erwartenden Festplattenplatzbedarf keine hinreichend genaue Abschätzung angeben.

Platzbedarf für gelöschte Dokumente

In ELO als gelöscht markierte Dokumente sind vollwertige Dokumente in der ELO iSearch. Sie müssen über die API such- und auffindbar sein, wenn im Client mit der Option für inklusive gelöschte Dokumente oder *nur gelöschte* Dokumente gesucht wird. Deshalb belegen gelöscht markierte Dokumente weiterhin Platz in den Caches der Elasticsearch im Java-Heap und im Filesystem-Cache (RAM). Auch in Bezug auf den belegten Plattenplatz sind es vollwertige Dokumente.

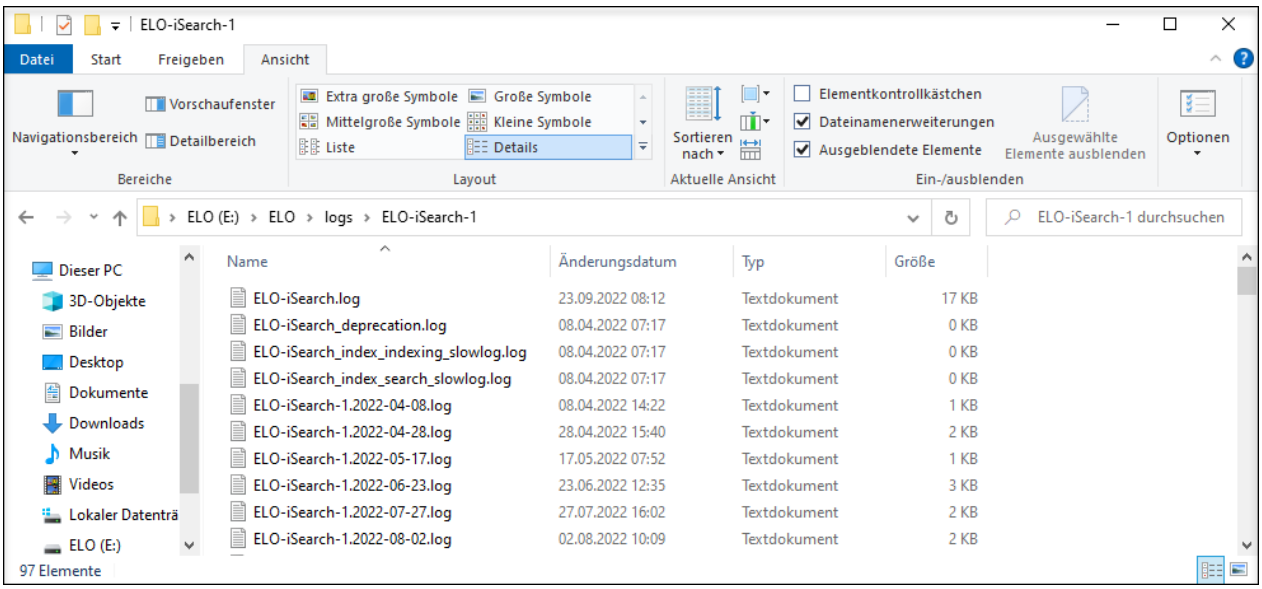
Final gelöschte Dokumente können dagegen nicht gesucht bzw. gefunden werden. Sie belegen keinen Platz mehr im Java-Heap oder Filesystem-Cache (RAM). Sie belegen allerdings weiterhin Festplattenplatz im Index, da sie nach internen Regeln der Elasticsearch bzw. Lucene gelöscht markiert und ggf. später bei Segment-Merge-Vorgängen nach und nach entfernt werden. Wenn dagegen ein Re-Index stattfindet, werden alle Indexe gelöscht und somit auch alle Dokumente in der ELO iSearch.

Betrieb

Für den Betrieb der Elasticsearch gibt es zwei wichtige Teilbereiche:

- Log-Dateien und deren Inhalt
- Erstellen von Auswertungen

Log-Datei



Standardmäßig liegen die Elasticsearch-Log-Dateien im ELO Installationsverzeichnis unterhalb des *logs*-Verzeichnisses.

Relevant ist hier die Log-Datei *ELO.log*, bzw. `ELO-<datum>.log` für ältere Log-Dateien.

Im Einzelfall kann auch ein Blick in die Log-Datei des Indexservers (`ix-<Repository>.log`) sinnvoll sein. Für ELO 11 und ELO 12 können Sie z. B. in der Log-Datei nach dem Begriff `queryTerm` suchen. Beispielhaft im Folgenden der Output für eine Volltextsuche nach dem Begriff `Test`:

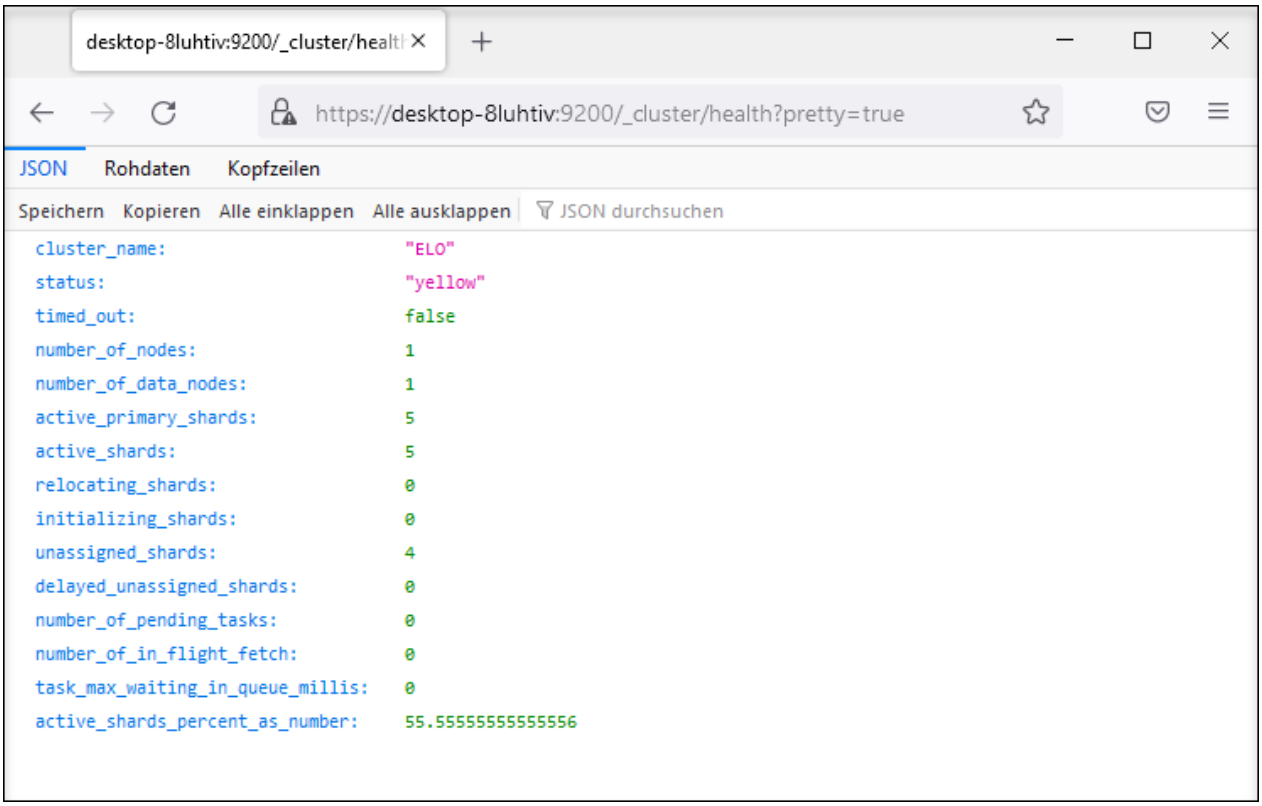
```
22:51:34,723 eloix-find-40 eloix-find-40 INFO (ElasticClient.java:183) - find(searchId=[
```

Auswertungen

Diverse Status der Elasticsearch können Sie per Browser abfragen. Diese können Sie dann gegebenenfalls auch für Monitoring-Lösungen (wie z. B. Nagios) verwenden. Zur Anmeldung ist jeweils ein administratives Tomcat-Konto notwendig. Ausgegeben wird jeweils eine JSON-Datei,

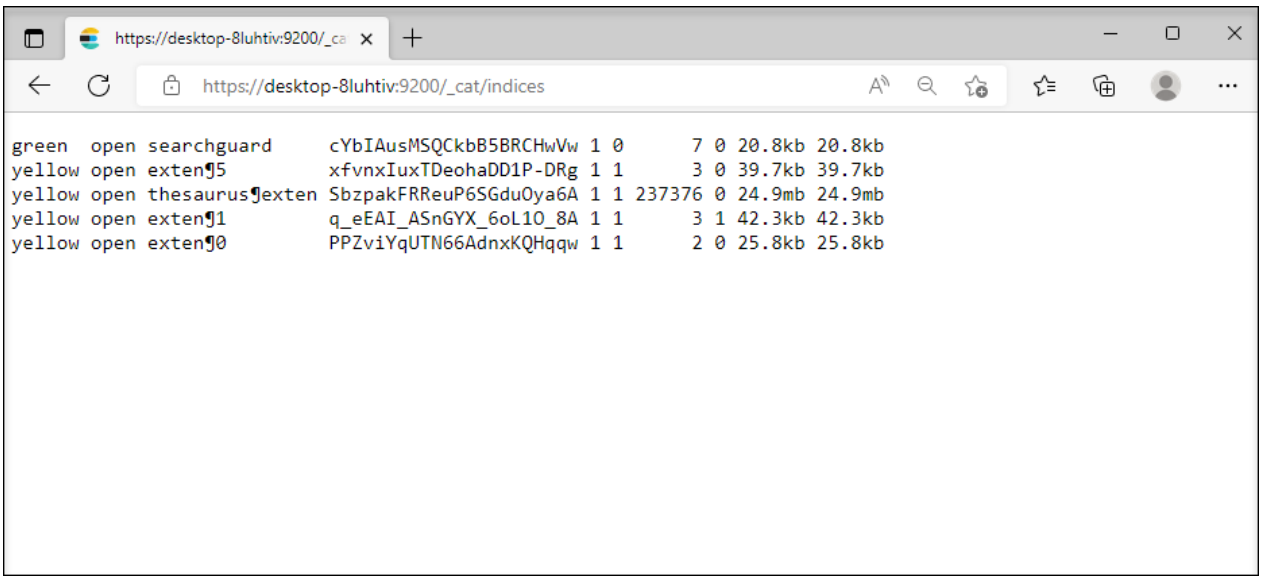
die Sie sich wie in dem folgenden Screenshot mit dem Browser Mozilla Firefox auch direkt anzeigen lassen können.

Die URL für *Health* lautet: `https://<server>:9200/_cluster/health?pretty=true`

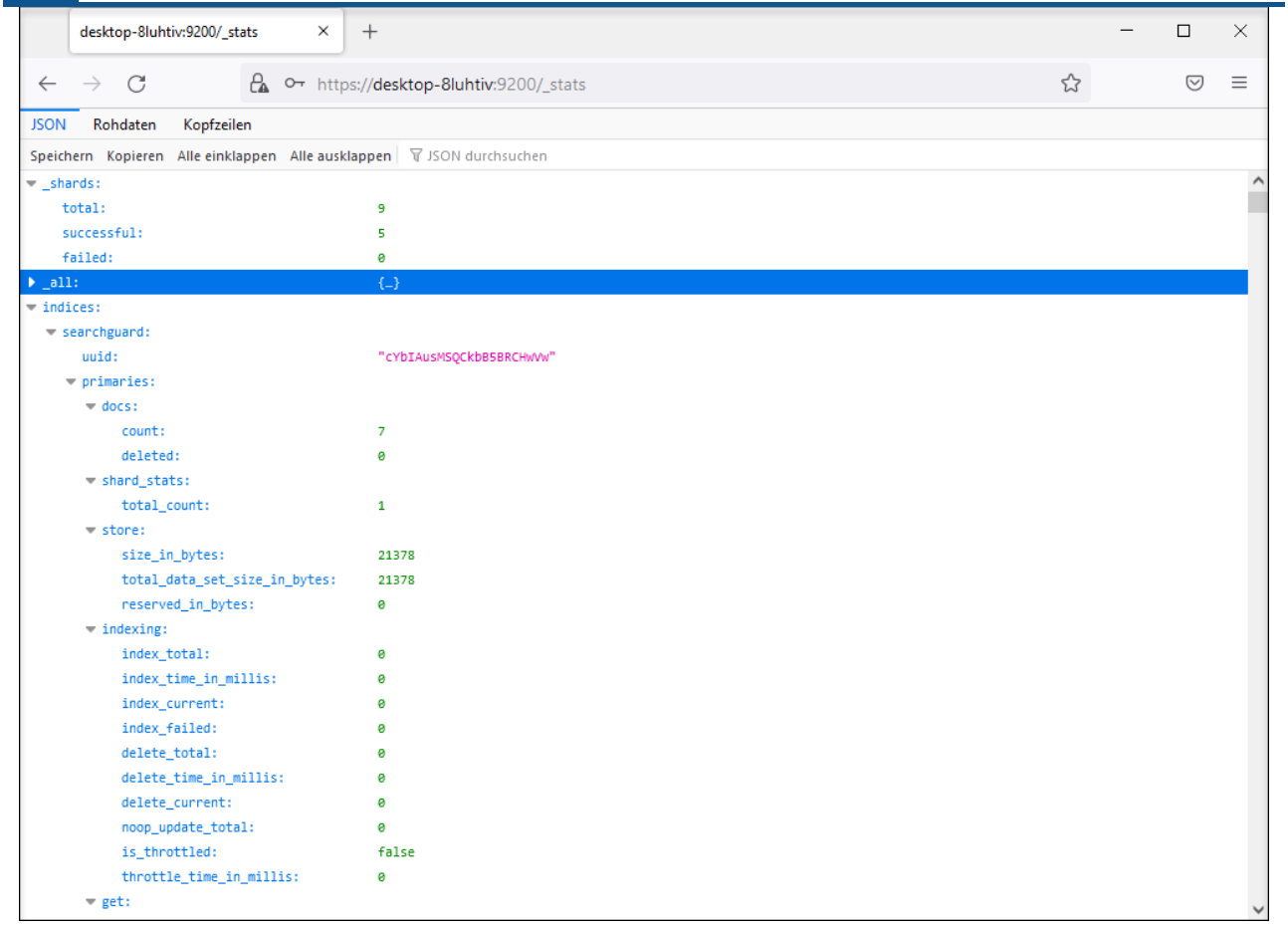


Der Status `yellow` bedeutet, dass der Index funktionsfähig ist, die Replica-Shards jedoch nicht alle verfügbar sind. In einer Einzelserverumgebung, bei der die Standardeinstellung für Replicas (1) genutzt wird, ist der Status `yellow` normal.

Die URL für Indexverzeichnis-Namen lautet: `https://<server>:<port>/_cat/indices`



Für jede Maske ist ein Index vorhanden, zu erkennen an der zugehörigen ID.



Die URL für Statistikdaten lautet: `https://\<server\>:9200/_stats`.

Hiermit können Sie prüfen, ob ein neues Dokument in den Index aufgenommen wurde. Nach Neuaufnahme eines Dokuments zeigt die Abfrage in der ersten Zeile nach `count` einen um 1 erhöhten Wert an.

In derselben Form sind weitere Abfragen möglich. Diese sind in der Online-Dokumentation der Elasticsearch beschrieben.

Backup-Möglichkeiten und Backup-Strategien

Die Daten der iSearch lassen sich zu 100 % aus der Datenbank über den Indexserver wiederherstellen. Die Entscheidung, ob ein Backup des Elasticsearch-Index angelegt wird, ist demnach von folgenden Faktoren abhängig:

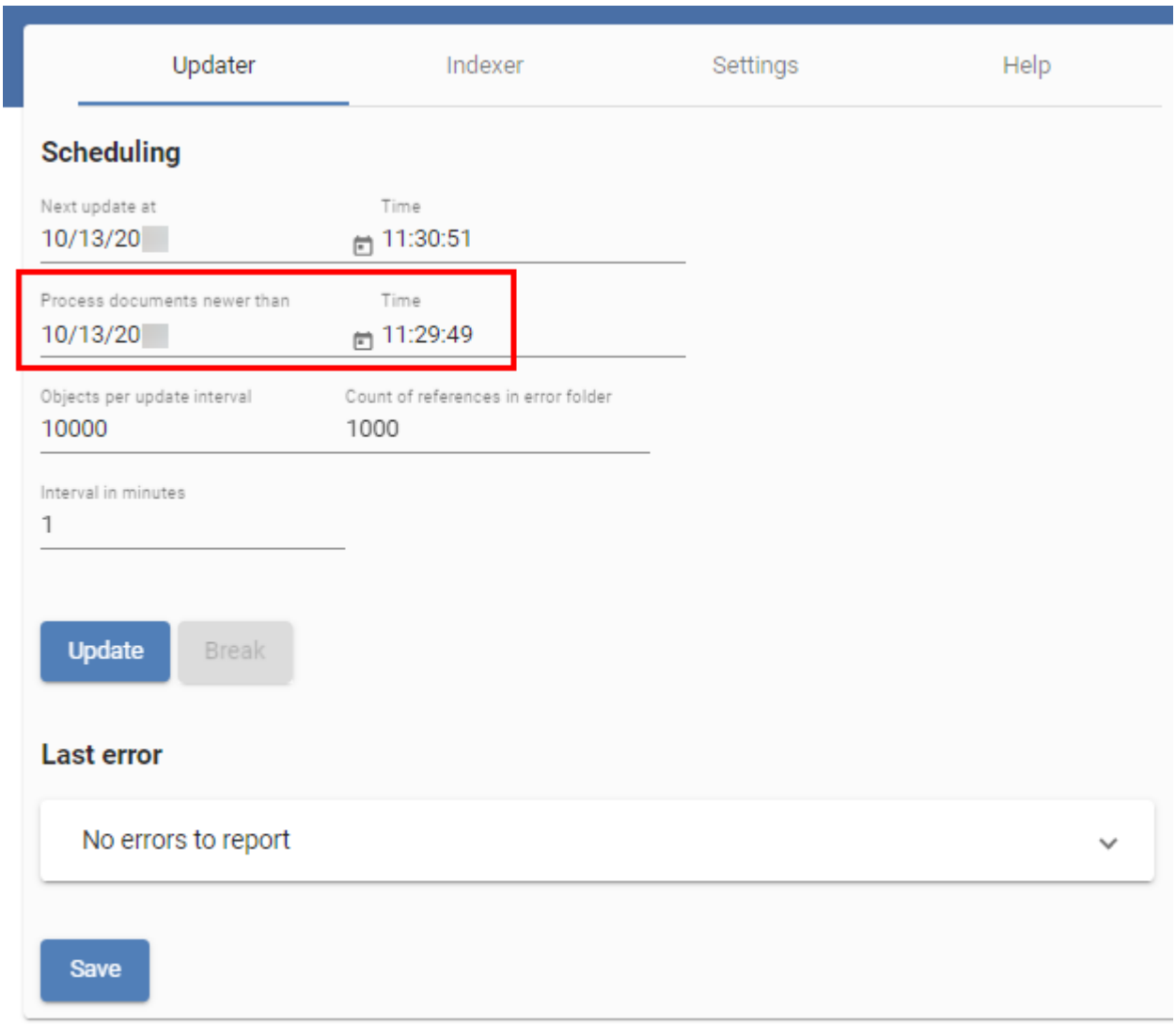
- Welche Downtime der iSearch ist für die Nutzer zumutbar?

Wenn die Downtime möglichst gering sein soll, ist eine Backup-Strategie notwendig, bei der der vorher gesicherte iSearch-Index innerhalb kurzer Zeit in das Produktivsystem gespielt werden kann.

- Wie lange dauert es, den iSearch-Index wiederherzustellen?

Die Dauer der initialen Re-Indexierung kann als Anhaltspunkt für folgende Re-Indexierungen genommen werden. Es muss entschieden werden, ob diese Dauer mit der Downtime für die Nutzer vereinbar ist.

Zum Erstellen eines Backups gibt es zwei Möglichkeiten. Bei beiden Lösungen müssen Sie beachten, dass die Indizes zwischen unterschiedlichen Indexserver-Versionen nicht unbedingt kompatibel sind. Im Anschluss an das Einspielen des Backups ist es notwendig, die Dokumente, die sich seit dem Erstellen des Backups geändert haben, erneut zu indexieren. Dies lässt sich über den Updater-Prozess der iSearch realisieren.



Der Zeitpunkt für *Process documents newer than* wird auf den Zeitpunkt gesetzt, zu dem das Backup erstellt wurde. Demnach ist es sinnvoll, z. B. täglich oder wöchentlich ein Backup anzulegen, damit die Dauer zum Indexieren der geänderten Dokumente nicht zu lange ist.

Zudem ist zu beachten, dass der Index mit dem Namen *searchguard* benötigt wird, um die Verschlüsselung mittels SSL über das *Search Guard-Plugin* zu gewährleisten. Existiert dieser Index nicht, so kann sich auch der Indexserver nicht mit der Elasticsearch verbinden. Haben Sie den *searchguard-Index* aus Versehen gelöscht, können Sie ihn durch erneutes Durchführen des Server-Setups erzeugen. Außerdem ist der *searchguard-Index* abhängig vom gewählten administrativen Tomcat-Konto, sodass der Index bei unterschiedlichen Konten oder Passwörtern nicht einfach kopiert werden kann, sondern durch das Server-Setup erzeugt werden muss.

Bestehenden Index kopieren

Den bestehenden Index können Sie aus dem Dateisystem kopieren und zu einem späteren Zeitpunkt wieder einspielen. Dazu kopieren Sie den kompletten Ordner unter *<Installationsverzeichnis>/data/<...><Servername>/index* als Backup. Sollte das Backup benötigt werden, ersetzen Sie den Inhalt des oben genannten Ordners mit dem Backup.

Den iSearch-Dienst sollten Sie vor dem Kopieren stoppen, um Zugriffsfehler zu vermeiden.

Snapshot mittels Elasticsearch erstellen

Elasticsearch selbst bietet die Möglichkeit, sogenannte Snapshots der bestehenden Indizes anzulegen. Werden regelmäßig Snapshots erzeugt, wird nicht jedes Mal eine komplette Kopie erzeugt, sondern nur die geänderten Dateien gespeichert und so der benötigte Speicherplatz reduziert.

Die Anleitung dazu finden Sie hier:

https://www.elastic.co/guide/en/elasticsearch/reference/7.15/snapshot-restore.html

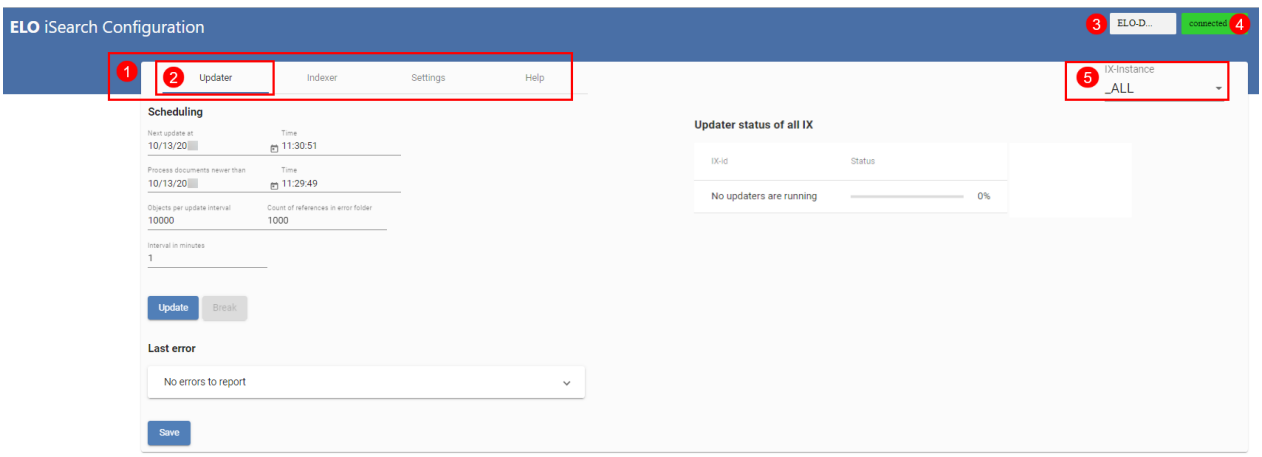
> **Beachten Sie**
>
> Um einen Snapshot wiederherstellen zu können, müssen Sie der Konfigurationsdatei der Elasticsearch (`<Installationsverzeichnis>/config/elastic/<Servername>/elasticsearch.yml`) folgende Zeile hinzufügen:
> `searchguard.enable_snapshot_restore_privilege: true`

Tipps und Tricks

ELO iSearch-Konfiguration

Über die Konfigurationsseite des Indexservers können Sie die *ELO iSearch Configuration* aufrufen, die ab ELO 11 zur Verfügung steht:

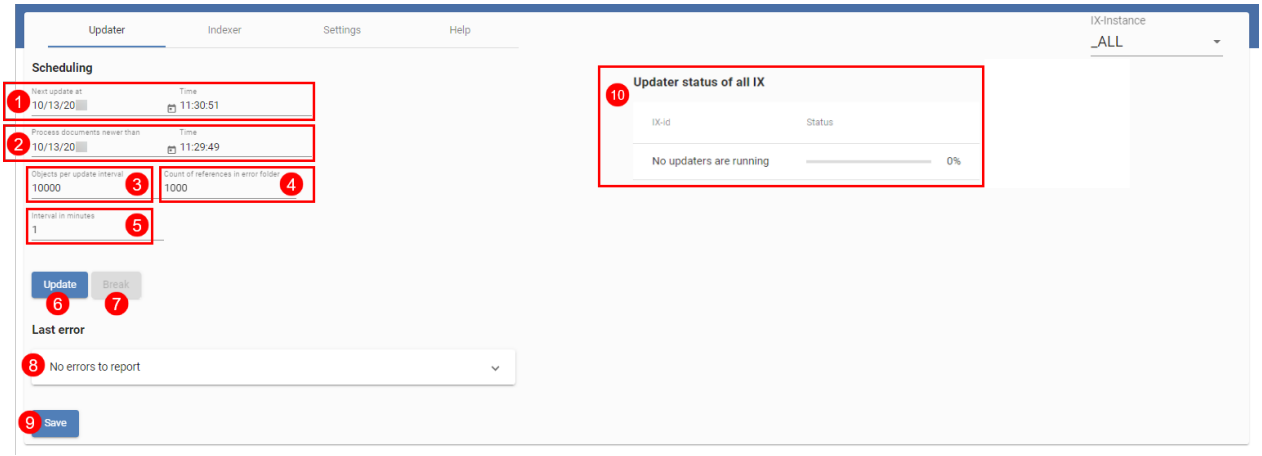
```
https://<server>:<Port>:/ix-<Repository>/manager/esconfig/#/isearchConfig
```



In der obigen Abbildung sehen Sie die Konfiguration der ELO iSearch. Sie können über die ELO iSearch Configuration die verschiedenen Funktionalitäten der iSearch aufrufen (1), der aktive Tab ist mit einem blauen Unterstrich markiert (2). Oben rechts (3) sehen Sie den Namen der ELOix-Instanz, über die die ELO iSearch Configuration aufgerufen wurde, deren Verbindungsstatus zur iSearch (4) ist *connected* bzw. *disconnected*.

Ab ELO 11 werden die Einstellungen für die iSearch in der Tabelle eloixopt gespeichert und sind demnach, wie auch die Indexserveroptionen, instanzspezifisch, bzw. für alle Instanzen gespeichert (*_ALL*). In der Tabelle haben alle iSearch-Einstellungen das Präfix *esearch*. Über die ELO iSearch Configuration können Sie die Einstellungen ebenso für alle Instanzen (*_ALL*) oder instanzspezifisch anzeigen lassen und bearbeiten (5).

Updater



In der obigen Abbildung sehen Sie die Übersicht über den Updater. Diesen benötigen Sie für die regelmäßige Aktualisierung der iSearch. Er sorgt zudem dafür, dass neue und geänderte Objekte an die iSearch gesendet werden. In der obersten Zeile (1) sehen Sie, wann der Updater das nächste Mal startet und in der nächsten Zeile (2), ab welchem Zeitpunkt Dokumente verarbeitet werden. Beide Werte können Sie einstellen. In der Standardkonfiguration läuft der Updater jede Minute (5) und verarbeitet maximal 1000 Objekte (3) in der kleinsten Zeiteinheit (1 Minute). Beachten Sie, dass die Anzahl der ausgewählten Objekte zeitgleich im Speicher des ELOix gehalten wird, den Wert sollten Sie daher nur mit Bedacht anpassen.

Kommt es zu Problemen beim Indexieren eines Dokumentes, wird eine Referenz auf das Dokument in ELO unter folgendem Pfad angelegt:

```
Administration // Fulltext Configuration // Update Errors
```

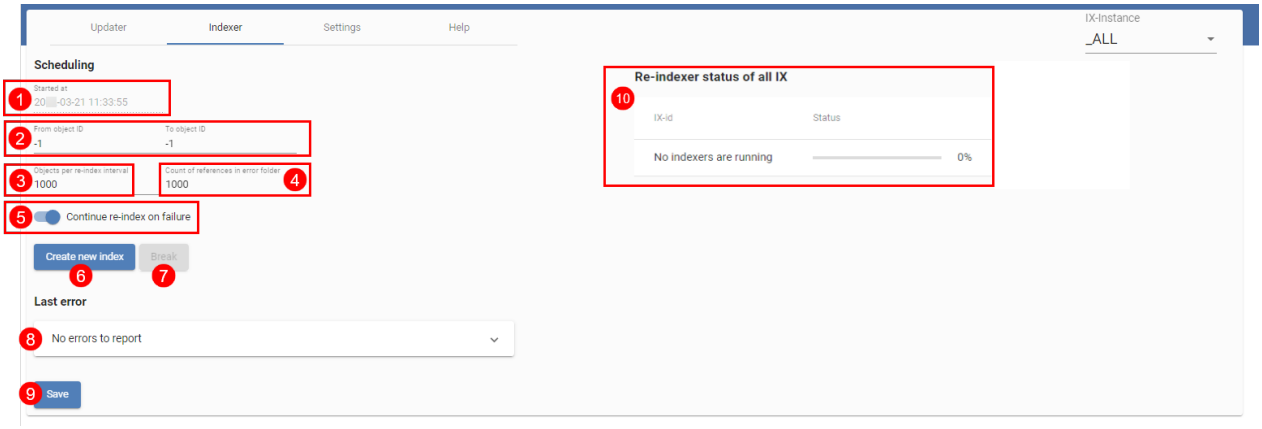
Dadurch können Sie leicht kontrollieren, ob es zu Fehlern kam und dafür sorgen, dass die Dokumente erneut an die iSearch gesendet werden, indem z. B. der Dialog *Metadaten* geöffnet wird. Die maximale Anzahl an Dokumenten in diesem Ordner stellen Sie in (4) ein. Stellen Sie `0` ein, werden keine Referenzen angelegt.

Nehmen Sie Änderungen vor, müssen Sie diese mit *Save* speichern (9).

Eine Übersicht über die laufenden Updater-Prozesse ist im Bereich *Updater status of all IX* (10) dargestellt. Einen Update-Prozess können Sie auch manuell über den Button *Update* (6) starten sowie über den Button *Break* (7) unterbrechen. Beachten Sie hierbei, dass sich das Starten und Unterbrechen immer auf die aktuelle ELOix-Instanz bezieht.

Der letzte Fehler des Updater-Prozesses wird in *Last Error* (8) angezeigt.

Indexer



Über den Indexer (obige Abbildung) nehmen Sie die initiale Befüllung der iSearch mit den Objekten aus ELO. Dazu werden alle vorhandenen Objekte an die iSearch gesendet. Gestartet wird mit den Objekten mit der höchsten ID, indexiert wird absteigend. Dadurch sind neue Objekte früher auffindbar, sofern der Index sofort produktiv verwendet wird.

In der obersten Zeile (1) ist der Startzeitpunkt des letzten Re-Index-Prozesses zu sehen. Über (2) und (3) wählen Sie die Objekt-IDs aus, die indexiert werden. Möchten Sie einen kompletten Re-Index durchführen, müssen Sie beide Werte auf `-1` setzen. Dies führt auch dazu, dass der bestehende Index gelöscht wird. Möchten Sie jedoch nur bestimmte Objekte erneut einem bestehenden Index hinzufügen, können Sie dies auch über die IDs steuern.

Die Einstellungen für die Objekte pro Indexierungsintervall (3) und Referenzen in ELO (4) entsprechen denen des Updaters. Die Referenzen werden im Ordner unter folgendem Pfad abgelegt:

```
Administration // Fulltext Configuration // Re-Index Errors
```

Dieser Wert ist der Gleiche, sowohl für den Updater als auch für den Indexer.

Die Einstellung *Continue Reindex on Failure* (5) gibt an, ob im Falle eines Fehlers der Re-Index-Prozess unterbrochen oder fortgesetzt werden soll. Möchten Sie ihn fortsetzen, ist ein regelmäßiges Kontrollieren der Log-Ausgaben unablässig.

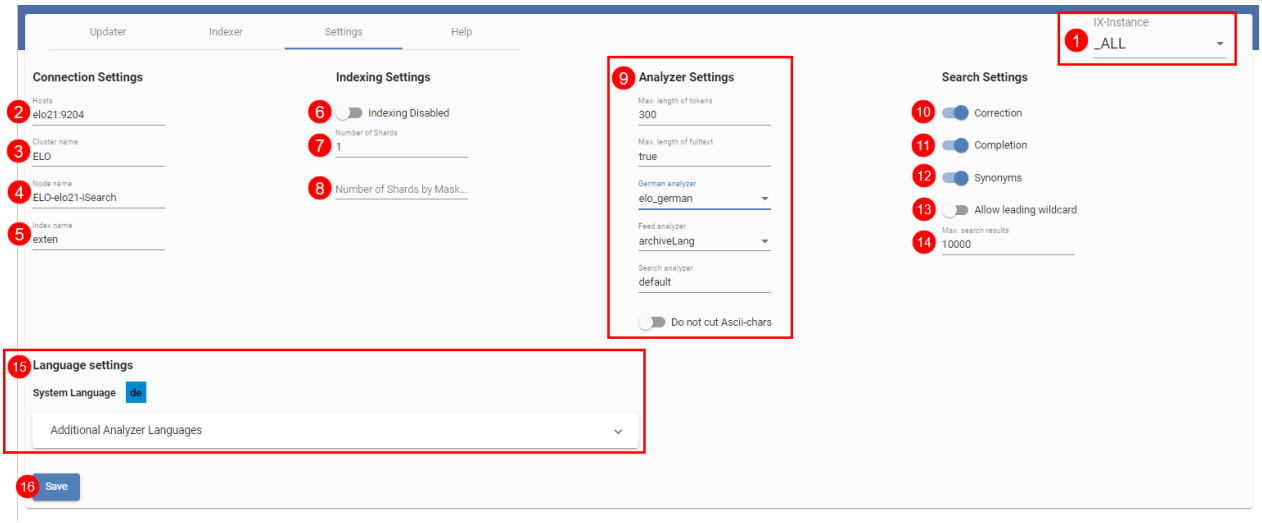
Der letzte Fehler des Re-Index-Prozesses wird in *Last Error* (8) angezeigt.

Die Einstellungen müssen Sie über den Button *Save* (9) speichern.

Ein Re-Index kann gestartet (wenn beide Objekt-IDs auf `-1` gesetzt sind) bzw. fortgesetzt (6) sowie unterbrochen (7) werden. Wie auch beim Updater müssen Sie beachten, dass dies sich immer auf die ELOix-Instanz bezieht, mit der die Config-Seite aufgerufen wurde.

Eine Übersicht über aktive Re-Index-Prozesse ist in (10) dargestellt.

Settings



Eine Übersicht über die weiteren Einstellungsmöglichkeiten der iSearch ist in der oberen Abbildung dargestellt. In *IX-Instance* (1) wählen Sie aus, ob die Einstellungen für alle Instanzen oder instanzspezifisch vorgenommen werden sollen (siehe auch die nächste Abbildung *ELO iSearch Configuration, Tab 'Setting', Einstellungen instanzspezifisch vornehmen*).

Die Details zum Verbindungsaufbau zwischen ELOix und Elasticsearch werden in (2), (3) und (4) angegeben. Der Name des Elasticsearch-Index (5) entspricht im Standard dem Namen des Repositorys in Kleinbuchstaben.

Des Weiteren können Sie angeben, ob ein ELOix überhaupt zum Senden von Objekten an die iSearch verwendet werden soll (6). Ist diese Einstellung aktiv, ist die Volltextsuche über die iSearch trotzdem noch aktiv.

Die Anzahl der Shards bzw. Shards pro Maske wird mit (7) und (8) eingestellt. In Kapitel Festlegung der Shard-Anzahl pro Index wird näher auf diese Einstellungen eingegangen.

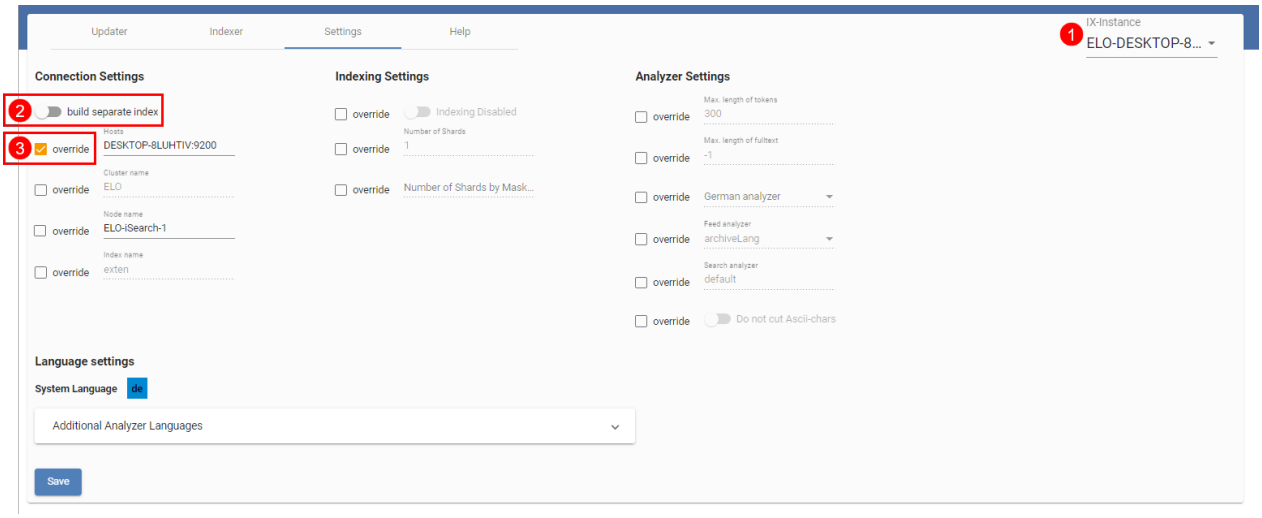
Die Analyzer-Einstellungen (9) beziehen sich darauf, wie Text verarbeitet wird, der an die Elasticsearch gesendet wird. Im Normalfall verwenden Sie die vorhandenen Voreinstellungen.

Die Einstellungen (10) bis (14) beziehen sich auf die Suche. Es ist einstellbar, ob Korrekturvorschläge (10) oder Synonyme (12) zur Suchanfrage zurückgegeben werden und ob die Autovervollständigung (11) bei der Sucheingabe aktiv ist. Des Weiteren können Sie die Suche mit einem Wildcard-Zeichen am Anfang des Suchwortes aktivieren (13). Dies ist standardmäßig deaktiviert, da eine Suche mit führendem Wildcard-Zeichen die Elasticsearch stark auslasten kann. Außerdem können Sie die maximalen Suchergebnisse einstellen (14), die über die ELOix-API-Aufrufe `findFirstSords()` und `findNextSords()` zurückgegeben werden.

Die Spracheinstellungen der iSearch können Sie mittels *Language Settings* (15) bearbeiten. Die System- bzw. Repository-Sprache wird über das Server Setup eingestellt und kann nicht bearbeitet werden. Je nach verfügbaren Lizenzen können Sie weitere Sprachen auswählen. Die Spracheinstellungen der iSearch können Sie mittels *Language Settings* (15) bearbeiten. Die System- bzw. Repository-Sprache wird über das ELO Server Setup eingestellt und kann nicht bearbeitet werden. Je nach verfügbaren Lizenzen können Sie weitere Sprachen auswählen. Die eingestellten Sprachen bestimmen bei der Indizierung, ob für ein Dokument mit einer erkannten Sprache spezielle sprachspezifische Analyzer-Schritte bei der Erzeugung der Tokens verwendet

werden oder nicht. Beispielsweise kann für die Sprache *deutsch* ein Decompounder-Filter benutzt werden, um zusammengesetzte Wörter zu zerlegen. Das führt dazu, dass auch nicht zusammengesetzte Teilwörter gefunden werden können. Für die Sprache *englisch* können z. B. typische Stoppwörter entfernt werden. Wenn während der Indizierung für ein Dokument eine Sprache erkannt wird, die nicht in den eingestellten Sprachen enthalten ist, wird das Dokument trotzdem indiziert, allerdings nur mit den Analyzer-Schritten, die nicht sprachspezifisch sind (Fallback-Analyzer).

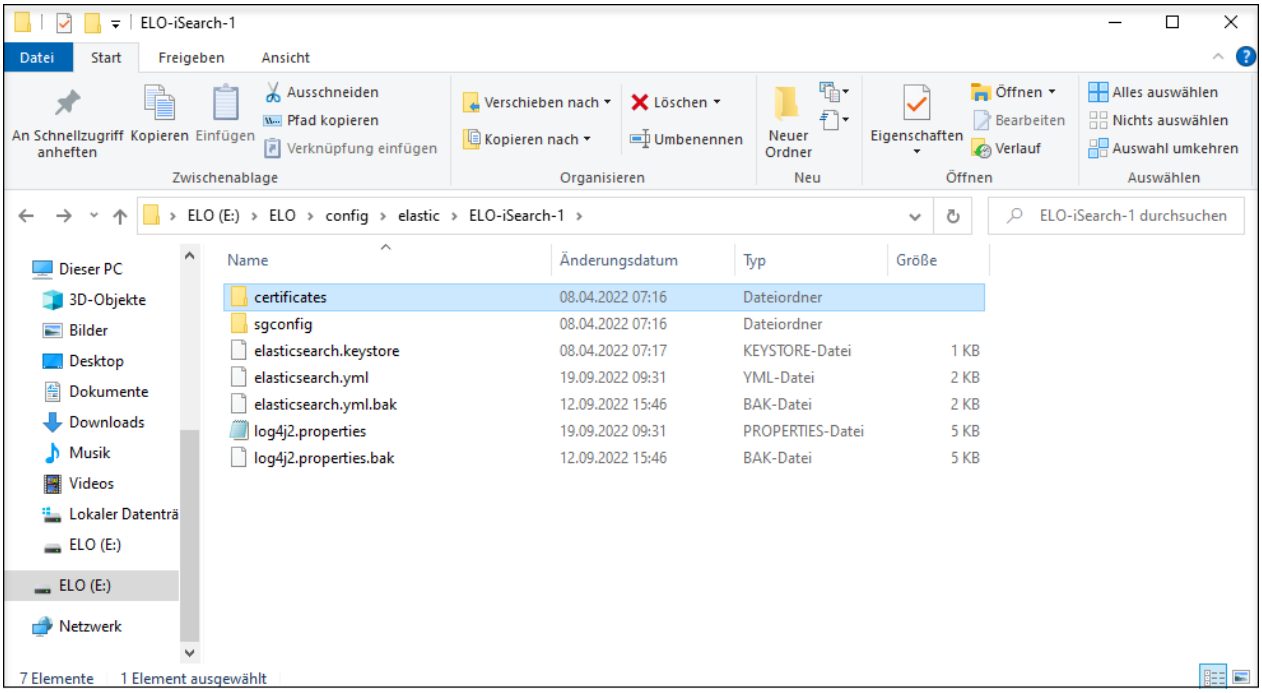
Alle Einstellungen müssen Sie über den Button *Save* speichern (16). Beachten Sie, dass die Einstellungen (5), (7), (8), (9) und (15) einen Re-Index benötigen, damit sie für alle Daten gelten und dies nicht zu unterschiedlichen Daten in der iSearch führt.



Die Abbildung zeigt beispielhaft die Einstellungsseite für die Instanz ELO-DESKTOP-8LUHTIV (1). Alle Einstellungen, bei denen die Override-Checkbox aktiv ist (3), überschreiben die Einstellung, die für alle Instanzen vorgenommen wird.

Die Einstellung *build separate index* (2) können Sie nur instanzspezifisch vornehmen. Ist diese Einstellung aktiv, wird mit diesem Indexserver ein separater Index aufgebaut – allerdings nur dann, wenn sich entweder die Verbindungseinstellungen, von denen für die anderen Instanzen unterscheiden oder der Indexname unterschiedlich ist. Im Kapitel Indexaufbau im Hintergrund wird dies näher erläutert.

Dienstekonfiguration der Elasticsearch



Die Konfiguration des Elasticsearch-Dienstes erfolgt über mehrere Dateien und Verzeichnisse.

Im Einzelnen haben die Dateien und Verzeichnisse folgende Inhalte:

- *elasticsearch.yml*: Zentrale Konfigurationsdatei der Elasticsearch. Detaillierte Informationen lesen Sie unter folgendem Link: https://www.elastic.co/guide/en/elasticsearch/reference/current/important-settings.html

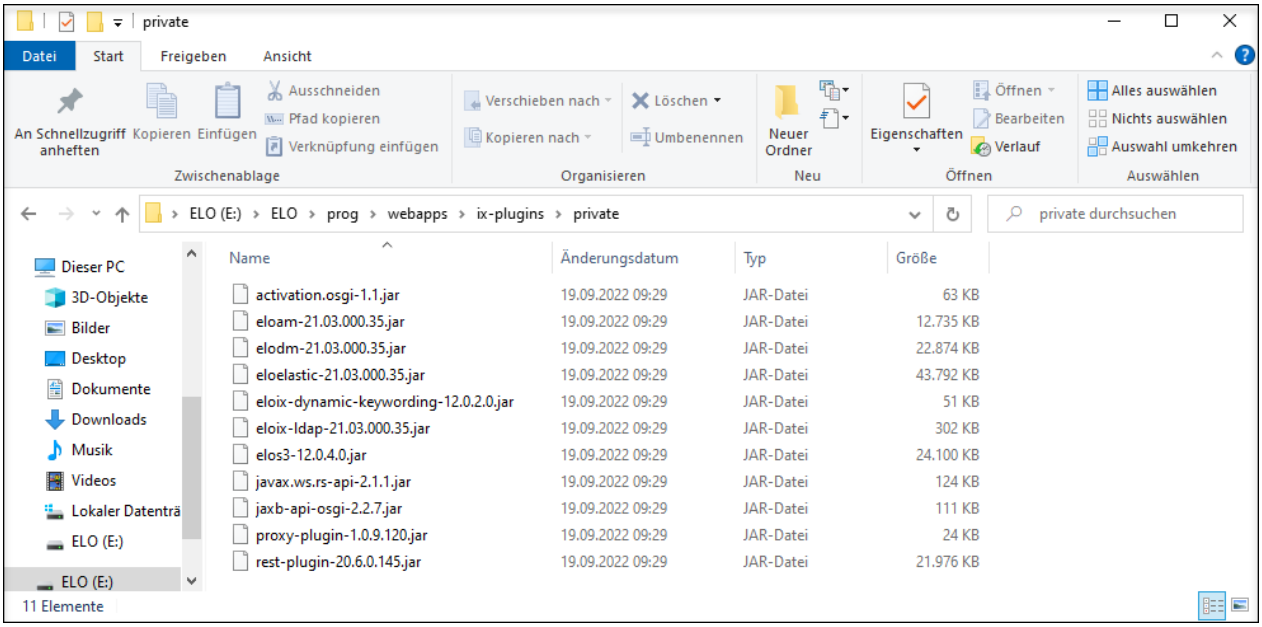
  > **Beachten Sie**
  >
  > Fügen Sie in der Konfigurationsdatei *elasticsearch.yml* außer auf Anweisung des ELO Supports keine Optionen hinzu.

- log4j.properties: Logging-Einstellungen
- sgconfig: Verzeichnis mit Konfigurationen des *Search Guard Plugins*
- certificates: Key- und Truststore für die SSL-Kommunikation mit dem ELO Indexserver

SSL-Kommunikation

Ab der ELO ECM Suite 10.2 kommunizieren die ELO Indexserver verschlüsselt mit der Elasticsearch. Dazu werden vom ELO Server Setup bei der Installation *selfsigned Zertifikate* und entsprechende Key- und Truststores angelegt. Die Key- und Truststores werden unterhalb der jeweiligen config-Verzeichnisse (Indexserver und Elasticsearch) angelegt und können mit Java-Standardmitteln (z. B. *keytool.exe*) für die Prüfung der Zertifikate geöffnet werden. Verwenden Sie das Passwort des administrativen Tomcat-Kontos.

Indexserver-Plugins



Ab ELO 11 ist es möglich, den Indexserver durch sogenannte OSGi-Plugins zu erweitern. Beispielsweise befindet sich die Elasticsearch-Anbindung selbst in einem solchen Plugin. Mit dieser Aufteilung ist es zukünftig möglich, mit einer älteren Indexserver-Version auf einen neueren Elasticsearch-Index zuzugreifen. Für jeden Indexserver wird ein Plugin-Verzeichnis konfiguriert.

Festlegung der Shard-Anzahl pro Index

Ab ELO 11 wird pro Maske ein Index angelegt. In der Standardeinstellung besteht jeder Index aus einem Shard. Durch die Aufteilung in einen Index pro Maske wird auch im Cluster schon ein Performance-Gewinn erreicht, da die Indizes auf mehrere Nodes aufgeteilt werden.

Zusätzlich ist es sinnvoll, besonders große Indizes in mehrere Shards aufzuteilen. Der Vorteil liegt darin, dass Suchanfragen dann parallel auf allen Shards ausgeführt werden. Allerdings werden maximal so viele Suchanfragen innerhalb eines Nodes parallel ausgeführt wie Prozessoren vorhanden sind.

Die Anzahl an Shards richtet sich demnach nach den enthaltenen Daten. Elasticsearch empfiehlt eine Shard-Größe von 20 bis 40 GB. Zu kleine Shards führen zu einer Überlastung in der Kommunikation zwischen den Shards, da die Daten zwischen den Shards synchronisiert werden müssen. Masken, die sehr viele Sords mit Volltext enthalten, sollten demnach möglichst in mehrere Shards aufgeteilt werden.

Die Anzahl an Shards pro Maske kann über die ELO iSearch Configuration vorgenommen werden, erfordert allerdings einen Indexaufbau (Re-Index) aller Daten.

Festlegung der Replica-Anzahl pro Index

Replicas sind gewissermaßen Kopien eines Index. Es kann jeder beliebige Wert gewählt werden, auch 0. Für jeden Shard wird ein Primary-Shard erzeugt, sowie im Cluster die angegebene Anzahl an Replica-Shards. Primary-Shards und Replica-Shards werden immer auf unterschiedlichen Nodes angelegt, um im Falle eines Ausfalls auch weiterhin die Daten

bereitstellen zu können. Fällt ein Node aus, werden die Primary-Shards, die bisher auf diesem Node allokiert waren, auf einem anderen Node allokiert. Die Replica-Shards werden auf andere Nodes verteilt, um weiterhin die bestmögliche Ausfallsicherheit und Performance zu gewährleisten.

Der weitere Vorteil von Replica-Shards ist die Parallelisierung von Suchanfragen: Die Elasticsearch verteilt Suchanfragen auf unterschiedliche Primary-Shards und Replica-Shards, um die Last der Suchanfragen zu verteilen.

Der Health-Zustand eines Indizes ist abhängig davon, welche Shards verfügbar sind: Ist der Primary-Shard vorhanden, ist der Zustand *gelb*, sind zusätzlich alle Replica-Shards verfügbar, ist der Zustand *grün*.

Die Menge der Replica-Shards sollten Sie an die Anzahl der Nodes im Cluster anpassen, sodass sich die Shards sinnvoll auf mehrere Nodes verteilen können, um Ausfallsicherheit zu gewährleisten. Andererseits benötigt jede Replica genauso viel Speicherplatz wie die Primary-Shard.

Die Standardeinstellung für die Anzahl an Replicas ist 1. So wird in einem Cluster von mehr als einer Instanz jeder Shard einmal repliziert, ist also insgesamt zweimal vorhanden. Eine Änderung an der Anzahl der Replicas erfordert keine Re-Indexierung.

Um eine Re-Indexierung in einem Cluster zu beschleunigen, können Sie für die Dauer der Re-Indexierung die Anzahl an Replicas auf `0` setzen und nach dem Beenden erhöhen. Ist die Anzahl an Replicas > 0, werden die Daten sowohl im Primary-Shard als auch in jedem Replica-Shard zeitgleich indexiert. Das bedeutet, dass die Analyse und Umwandlung in Elasticsearch-Daten mehrmals parallel stattfinden. Wird die Anzahl an Replicas auf `0` gesetzt, werden die Daten während der Re-Indexierung nur in einem Node verarbeitet. Sobald die Anzahl an Replicas erhöht wird, werden die bereits indexierten Primary-Shards auf die anderen Nodes kopiert, was in der Regel deutlich schneller geht. Natürlich ist dadurch während der Re-Indexierung keine Ausfallsicherheit gewährleistet.

Die Einstellung ist bezogen auf einen Index und kann z. B. mit dem curl-Tool (https://curl.haxx.se/) exemplarisch folgendermaßen für die Maske mit der ID `42` des Repositorys *repository* auf `2` geändert werden:

```
curl -X PUT "localhost:9200/repository¶42/_settings" -H 'Content-Type: application/json'
{
    "index" : {
       "number_of_replicas" : 2
    }
}
'
```

Möchten Sie die Einstellung für alle Indizes ändern, können Sie die Angabe des Index weglassen und die URL `localhost:9200/_settings` verwenden.

Tokenisierung

Die Option *Tokenisierung deaktivieren* einer Feldvorlage schaltet die Suchregeln und -muster für dieses Feld aus. Es erscheinen dann in der Liste des Filters die exakten Werte der Metadaten.

Die Suche arbeitet für dieses Feld dann ebenfalls in einem exakten Modus. Teilbegriffe werden in diesem Fall nicht berücksichtigt.

Bei aktiven Suchregeln und -mustern erscheinen in der Liste des Filters nur einzelne Wörter, sowohl die Originale als auch die erzeugten Stammformen. Die Suchmöglichkeiten in diesem Filter entsprechen dann denen des Eingabefeldes der Suche.

Einbinden eigener Thesauri

Die Wörterbücher für den Thesaurus der ELO iSearch können Sie für ein Repository individuell anpassen. Die benötigten Dateien liegen in ELO unter folgendem Pfad:

`Administration // Fulltext Configuration // Thesaurus`

In diesem Pfad befindet sich für jede Sprache ein Unterverzeichnis mit dem Länderkürzel, in dem sich jeweils drei Dokumente befinden.

<Sprache>_thesaurus: Der von ELO ausgelieferte Thesaurus für diese Sprache. Diese Datei sollten Sie nicht ändern, damit sie bei einem Update gegen eine neuere Version ausgetauscht werden kann.

<Sprache>_add: In diesem Dokument können Sie zusätzliche Begriffe eintragen, welche als Synonyme in den Thesaurus aufgenommen werden sollen.

<Sprache>_stop: Über dieses Dokument können Sie einzelne Begriffe aus dem Thesaurus ausschließen, wenn diese Ihnen für Ihren Einsatz unpassend erscheinen.

> **Beachten Sie**
>
> Nach einer Änderung der Dokumente muss der ELO Indexserver neu gestartet werden, damit der Thesaurus anhand der geänderten Dateien neu aufgebaut werden kann.

Cluster-Betrieb

Wie Sie ELO in einem Cluster betreiben, lesen Sie in der Dokumentation zur ELO iSearch in ELO Server - Installation und Betrieb.

Empfehlungen bei Fehlermeldungen

Fehlermeldung enthält Text 'pending translog recovery'

Nach einem ungeplanten Neustart des ELO iSearch-Dienstes kann es im iSearch-Log zu Fehlermeldungen kommen, die den Text `pending translog recovery` enthalten.

Beispiel:

```
[2023-06-29T13:22:51,662][WARN ][o.e.i.f.SyncedFlushService] [ELO-iSearch] failed to flu
java.lang.IllegalStateException: [contelo¶2][0] flushes are disabled - pending translog
        at org.elasticsearch.index.engine.InternalEngine.ensureCanFlush(InternalEngine.j
```

Diese Meldungen können sehr zahlreich sein und erscheinen dann, wenn die ELO iSearch beendet wurde und vor dem Herunterfahren nicht mehr dazu kam, alle Transaktionen in den Index zu überführen (z. B. bei Hauptspeichermangel oder wenn die Speicherplatte voll ist). Beim Hochfahren versucht die Elasticsearch dann das Translog-File in den Index zu überführen. Das kann bei vielen Daten eine Stunde oder sogar mehrere Stunden dauern. Während dieser Zeit sind die Indexe nicht verfügbar.

Empfehlung:

Hier müssen Sie zunächst abwarten, bis das gesamte Translog in den Index überführt ist und diese Meldungen nicht mehr erscheinen. Danach sollten Sie untersuchen, welches ursprüngliche Problem zum Herunterfahren des ELO iSearch-Dienstes geführt hatte.

Vermeiden Sie es, eine Re-Indexierung über die Konfigurationsseite der ELO iSearch zu starten, während diese Meldungen erscheinen. In der Regel ist der Index nach Abschluss des Translog-Recovery wieder konsistent und erfordert keine Re-Indexierung.

Wenn während des Translog-Recovery das System hart heruntergefahren wird, kann es zu endgültig korrupten Indexdaten kommen, die durch das Translog-Recovery nicht mehr behebbar sind. Die Folge ist die Notwendigkeit einer Re-Indexierung oder sogar einer Neuinstallation des ELO iSearch-Dienstes.

Abbruch der Re-Indexierung wegen zu vieler Feed-Beiträge

Wenn es bei einem Re-Indexierungsvorgang der iSearch zu einem Abbruch kommt, kann in seltenen Fällen als Ursache eine extrem große Menge an Feed-Beiträgen infrage kommen. Dies kann hervorgerufen werden durch fehlerhafte Automatismen wie z. B. Importprozesse oder Scripte, die Dokumente ablegen und dabei ungewollt eine enorme Anzahl an Feed-Beiträgen ggf. mit Fehlermeldungen als Inhalt erzeugen.

Das lässt sich an zwei Stellen erkennen:

- Der ELO Indexserver bricht ab mit der Fehlermeldung: `java.lang.OutOfMemoryError: Java heap space`.
- Im Log des ELO Indexservers sind in Bezug auf die *feedaction*-Tabelle SQL-Exceptions mit `read timed out` zu sehen. Dabei sind zusätzlich zuvor Einträge mit einer hohen Anzahl an Feed-Beiträgen pro einzelnem Dokument zu sehen.

  Beispiel:

  ```
  19:20:58.529 INFO esin-8 esin-8 (FeedIndexer.java:108) - [821022] indexFeed=true: posts=#145239
  ```

Um festzustellen, ob zu viele Feed-Beiträge die Ursache für den Abbruch sind, können Sie mit einer Abfrage in der Datenbank ermitteln, zu welchen SORDs (d. h. ELO Dokumenten) eine sehr hohe Anzahl an Feed-Beiträgen existiert.

Beispiel einer SQL-Abfrage:

```sql
select count(*), objid
from documentfeed d
inner join feedaction f on d.feedguid=f.feedguid
inner join objekte o on d.objguid = o.objguid
group by objid
having count(f.feedguid) > 1000
order by 1 desc
```

Die Anzahl `1000` dient beispielhaft als eine Schwelle für viele Feed-Beiträge pro ELO Dokument. Wenn hier sehr hohe Werte (Zehntausende oder Millionen) in der ersten Ergebnisspalte zu sehen sind, müssen Sie die Ursache für diese Feed-Einträge finden und fehlerhafte Automatismen korrigieren.

> **Information**
>
> Bei kleinen ELO Systemen kann auch eine Gesamtmenge von vielen Millionen Feed-Beiträgen eine erhöhte Last oder einen erhöhten Speicherverbrauch des ELO

> Indexservers verursachen. Hier ist die Gesamtanzahl an Einträgen in der Tabelle *feedaction* relevant.

Bringen Sie das System nach der Korrektur von fehlerhaften Automatismen folgendermaßen wieder in einen arbeitsfähigen Zustand.

Vorgehen

1. Sichern Sie vorab die Datenbank über ein Backup.
2. Stoppen Sie alle ELOix-Instanzen für die Dauer der Korrektur.
3. Identifizieren Sie die *feedguid* zur betreffenden *objguid* in der Tabelle *documentfeed*.
4. Löschen Sie in der Tabelle *feedaction* die Zeilen mit der *feedguid*, die nicht mehr benötigt werden oder fehlerhaft sind.
5. Optional: Identifizieren Sie in der Tabelle *feedactionhist* (Änderungshistorie) die entsprechenden *actionguids* zur ermittelten *feedguid* und löschen Sie dann in der Tabelle *feedactionhist* die zu diesen *actionguids* passenden Zeilen. Diese wirken sich nicht direkt auf den Re-Index aus, sind allerdings überflüssig, wenn die Haupt-Einträge gelöscht sind.
6. Falls alle zugehörigen Zeilen in der Tabelle *feedaction* gelöscht wurden (d. h. falls alle diese Feed-Posts fehlerhaft oder ungewollt sind), löschen Sie in der Tabelle *documentfeed* die Zeile mit dieser *feedguid* (*feedaction.feedguid = documentfeed.feedguid*).

Alternativ:

Falls nur alte oder nicht mehr benötigte Feed-Beiträge bereinigt werden sollen, reicht es, wenn Sie die Schritte 1 bis 6 außer Schritt 3 durchführen. Dabei können z. B. alle noch benötigten Einträge in eine zusätzliche Tabelle kopiert werden und danach diese in *feedaction* umbenannt werden.

Fehlermeldung enthält Text 'disk usage exceeded flood-stage watermark'

Die Fehlermeldung `disk usage exceeded flood-stage watermark` im Log der ELO iSearch deutet darauf hin, dass auf mindestens einem Datenknoten ein kritischer Füllstand der Festplatte erreicht wurde. Sobald der *Flood-Stage Watermark* überschritten wird, schützt die Elasticsearch den Cluster vor einem vollständigen Speicherüberlauf. Dies geschieht standardmäßig bei ca. 95 % Festplattenbelegung.

Beispiel:

```
Zeile 829685: de.elo.ix.priv.plugin.esearch.exception.ESearchException: [ESEARCH:5524] f
Zeile 829686: at de.elo.elastic.ElasticClient.indexSord(ElasticClient.java:578)
Zeile 829693: Caused by: org.elasticsearch.ElasticsearchStatusException: Elasticsearch e
```

Empfehlungen:

1. Geben Sie Speicherplatz frei.
    - Alte Log-Dateien entfernen
    - Datenträger erweitern
2. Auf großen Festplatten kann ggf. genügend Speicherplatz für den Index der Elasticsearch zur Verfügung stehen und es kommt trotzdem weiterhin zur Anzeige dieser

Fehlermeldung. Teilen Sie in diesem Fall die Festplatte so auf, dass auf der geteilten Platte ausschließlich die Elasticsearch läuft. Wenden Sie sich an den ELO Support, falls die Probleme bestehen bleiben.

3. Starten Sie danach die ELO iSearch neu.

Powersuche

Die Powersuche der ELO iSearch bietet die Möglichkeit, komplexe Anfragen an die iSearch zu stellen und direkt auf deren Syntax zuzugreifen.

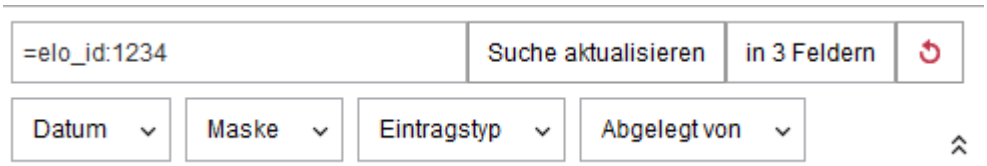
Die Powersuche der ELO iSearch erlaubt es Ihnen, direkt auf die zugrundeliegende Elasticsearch zuzugreifen. Dadurch bieten sich zwei neue Möglichkeiten: Zum einen kann auf Felder, die nicht über einen Filter auswählbar sind, direkt zugegriffen werden. Zum anderen können über die von Elasticsearch angebotene Syntax sehr komplexe Anfragen gestellt werden.

Der direkte Zugriff auf die internen Strukturen ist gleichzeitig auch der Nachteil der Powersuche, denn die Schreibweise ist sehr technisch und erfordert Wissen über die verwendeten internen Datenstrukturen. Dieser interne Aufbau kann sich zudem in kommenden ELO Versionen ändern, sodass vorhandene Suchabfragen nicht mehr richtig oder gar nicht mehr funktionieren.

> **Achtung**
>
> Die Powersuche sollte nur in besonderen, einzelnen Fällen benutzt werden, wenn die normalen Suchmöglichkeiten nicht ausreichen (z. B. bei administrativen Spezialanfragen).

Anwendung der Powersuche

Die Powersuche kann sowohl über den ELO Java Client als auch den ELO Web Client verwendet werden. Die Suchanfrage wird über die Suchzeile eingegeben und mit einem = ausgelöst, wie im folgenden Beispiel zu sehen ist.



Mit der Powersuche werden Anfragen über den ELO Indexserver direkt an die Elasticsearch geschickt. Der ELO Indexserver fügt eine Berechtigungsüberprüfung hinzu, sodass Sie trotz Powersuche immer nur Dokumente angezeigt bekommen, für die Sie eine Berechtigung haben.

Die Syntax der Powersuche entspricht der Syntax der *QueryStringQuery*, wie sie von Elasticsearch unter diesem Link dokumentiert wird:

https://www.elastic.co/guide/en/elasticsearch/reference/5.6/query-dsl-query-string-query.html#query-string-syntax

Für die Benutzung ist wichtig, dass Sie Sonderzeichen nun selbst escapen müssen. Dies ist notwendig, wenn ein Sonderzeichen wie ein normaler Buchstabe interpretiert werden soll und nicht in seiner besonderen Bedeutung, welche die Suche beeinflusst. Folgende Sonderzeichen

haben eine solche besondere Bedeutung in der *QueryStringQuery*-Syntax und müssen, wenn diese Bedeutung nicht gewünscht ist, escaped werden:

Zeichen	Bedeutung
+<wort>	<wort> muss im Dokument vorkommen
-<wort>	<wort> darf im Dokument nicht vorkommen (nur wenn davor ein Leerzeichen steht)
<wort1> && <wort2>	Beide Worte müssen im Dokument vorkommen
<wort1> und <wort2> getrennt durch zwei Pipe-Symbole	Mindestens eines der beiden Worte muss im Dokument vorkommen
> x < x	Größer/ kleiner als ...-Suche (funktioniert mit Zahlen oder mit Worten)
>= x <= x	Größer gleich/kleiner gleich als ...-Suche (funktioniert sowohl mit Zahlen als auch mit Worten)
=	Allein keine Bedeutung, am Anfang der Zeile: Powersuche
{x TO y}	Bereichssuche (ohne x und y)
[x TO y]	Bereichssuche (inklusive x und y)
<wort>^	Gewichtung des Wortes
"<wort> <wort>"	Phrasensuche
<wort>~	Fuzziness (unscharfe Suche)
*	Wildcard, beliebig viele Zeichen
?	Wildcard, exakte Anzahl an zu ersetzenden Zeichen
\	Escaped-Zeichen
/	Keine Bedeutung
<feldname>:<wort>	Suche in <feldname> nach <wort>

Feldnamen

Einzelne Felder der Elasticsearch können direkt abgefragt werden, sofern der Feldname bekannt ist. Eine Übersicht über die Feldnamen wird von ELO nicht bereitgestellt. Die Administration kann sie jedoch über die REST-Schnittstelle der Elasticsearch abfragen:

```
http://<host>:<port>/<Name des Repositorys in Kleinbuchstaben>/_mapping?pretty
```

Weitere Informationen zur Interpretation der Abfrage bietet Elasticsearch unter diesem Link: *https://www.elastic.co/guide/en/elasticsearch/reference/5.6/index.html*

> **Beachten Sie**
>
> ELO behält es sich vor, die Feldnamen zwischen den Versionen zu ändern. Daher ist es möglich, dass eine Anfrage mit einer Version funktioniert, mit der nächsten Version jedoch nicht mehr.

Beispiele für Abfragen

Oft ist es sinnvoll herauszufinden, ob ein bestimmtes Dokument über die iSearch auffindbar ist. Dies kann über folgende Suchanfrage herausgefunden werden:

```
=elo_id:<Objektid>
```

Voraussetzung für die Suche ist, dass Sie mit dem Recht *Alle Einträge sehen, Berechtigungen ignorieren* suchen. Falls dies zu keinem Ergebnis führt, ist das Dokument bisher nicht in die iSearch aufgenommen worden.

Die Suche nach `EL0[123` führt zu einem Syntaxfehler, da die eckige Klammer ein Sonderzeichen ist, das auch mit der Standardsuche für eine Bereichssuche genutzt werden kann. Um trotzdem nach diesem Zeichen zu suchen, kann die Powersuche verwendet werden, indem folgende Abfrage gestellt wird:

```
=EL0\[123
```

Lizenzbestimmungen bei der Verteilung von Serverprozessen

ELOprofessional-Systeme sind nur bedingt clusterfähig, sie erlauben ausschließlich den Hot-Standby-Betrieb der einzelnen Komponenten. ELOenterprise-Systeme sind dagegen voll clusterfähig.

Bei ELOprofessional dürfen die Serverprozesse nur innerhalb desselben Hosts, also auf einem Serversystem, verteilt werden. Bei ELOenterprise hingegen lassen sich die Prozesse auf mehrere Hosts (VMs oder physische Server) verteilen.

	ELO ECM Suite	
Funktionsbereiche	**ELO**enterprise	**ELO**professional
Betriebssystem	Plattformunabhängig[2]	Windows[1]
Architektur	SOA-konforme modulare Client-Server Architektur	SOA-konforme modulare Client-Server Architektur
Max. Anzahl Archive	unbegrenzt	20 pro Archivserver
Max. Anzahl Anwender	unbegrenzt	1000
Max. Anzahl Dokumente	2 Mrd. pro Archiv	100 Mio
Max. Archivebenen	254	254
Datenbank	PostgreSQL / MS SQL / Oracle / IBM DB2 [3]	PostgreSQL / MS SQL / Oracle / IBM DB2 [3]
Mandantenfähig	✓	•
Portalanbindung	✓	✓
Statuslog für zentrale Monitor-Systeme	✓	✓
Verteilbarkeit von Serverprozessen	✓	Single Host
Clusterfähigkeit	✓	Hot-Standby

Im Rahmen von ELOprofessional können somit die ELO Serverkomponenten (z. B. ELO Automation Service (ELOas)) auf mehreren Tomcat-Servern verteilt werden, solange sich diese auf einem Serversystem (VM oder physischer Server) befinden. Bei ELOenterprise ist es

hingegen erlaubt, die Tomcat-Server auf mehreren Serversystemen (VMs oder physischen Servern) zu installieren.

Glossar

Begriff	Erklärung
Escapen	Das Maskierungszeichen (Escape-Zeichen) ist ein bestimmtes Zeichen, das verhindert, dass das nachfolgende Zeichen vom ausführenden Programm als Funktionszeichen erkannt wird. Das Escapen ist also dann notwendig, wenn ein Sonderzeichen wie ein normaler Buchstabe interpretiert werden soll und nicht in seiner besonderen Bedeutung.
Garbage Collection (GC)	Die Garbage Collection , kurz GC (auch automatische Speicherbereinigung oder Freispeichersammlung genannt) bezeichnet in der Software- und Informationstechnik eine automatische Speicherverwaltung, die den Speicherbedarf eines Computerprogramms minimiert. Dabei wird zur Laufzeit versucht, nicht länger benötigte Speicherbereiche automatisch zu identifizieren, um diese dann freizugeben.(Quelle: https://de.wikipedia.org/wiki/Garbage_Collection, Stand, 29.03.2022)
Lucene	Apache Lucene ist eine hochperformante, frei verfügbare Programmbibliothek, die sich besonders gut zur Volltextsuche eignet. Lucene kann jede Art von Dokumenten indizieren. Dies können Texte sein, aber auch Zahlen, Datumswerte oder GPS-Positionen. Lucene ist die unter der Such- und Analytik-Engine Elasticsearch liegende Struktur zur Datenspeicherung.
NODE / NODES	NODE – engl. node bedeutet Knoten – NODES – Plural. Ein Elasticsearch-Server entspricht einem Node.
OCR	Optische Zeichenerkennung (englische Abkürzung OCR von *optical character recognition*) bezeichnet die automatische Texterkennung innerhalb von Bildern. Gescannte Dokumente werden durch die OCR in eine Textform umgewandelt.
OSGi (OSGi-Plugins)	Die OSGi Alliance (früher *Open Services Gateway initiative*) spezifiziert eine hardwareunabhängige dynamische Softwareplattform, die es erleichtert, Anwendungen und ihre Dienste per Komponentenmodell (*Bundle/Service*) zu modularisieren und zu verwalten (*Service Registry*). Die OSGi-Plattform setzt eine Java Virtual Machine (JVM) voraus und bietet darauf aufbauend das OSGi-Framework.(Quelle: https://de.wikipedia.org/wiki/OSGi, Stand, 29.03.2022)
REST	Representational State Transfer (abgekürzt REST , seltener auch ReST) bezeichnet ein Programmierparadigma für verteilte Systeme, insbesondere für Webservices. REST ist eine Abstraktion der Struktur und des Verhaltens des World Wide Web. REST hat das Ziel, einen Architekturstil zu schaffen, der die Anforderungen des modernen Web besser darstellt. Dabei unterscheidet sich REST vor allem in der Forderung nach einer einheitlichen Schnittstelle. Der Zweck von REST liegt schwerpunktmäßig auf der Maschine-zu-Maschine-Kommunikation.(Quelle: https://de.wikipedia.org/wiki/Representational_State_Transfer, Stand, 29.03.2022)
Replikation	Replikation bezeichnet die mehrfache Speicherung derselben Daten. In der Elasticsearch können Sie mit dieser Backup-Funktion einstellen, wie viele Replikate Sie von einem Shard haben möchten. Diese werden dann intelligent auf die Cluster-Nodes aufgeteilt. Somit wird die Suche auf mehrere Nodes verteilt und beim Ausfall eines Nodes sind Replikationen der Daten vorhanden, sodass die Suche ohne Datenverlust weiterhin verfügbar ist.

Begriff	Erklärung
Shard	Die Daten eines Elasticsearch-Suchindex werden in Shards aufgesplittet (wörtlich übersetzt *Scherbe*).
Sprach-Analyzer	Die Elasticsearch wird mit Analyzern der gängigsten Sprachen ausgeliefert. Diese Sprach-Analyzer führen im Wesentlichen vier Funktionen aus: Sie teilen den Text in einzelne Wörter auf, setzen die Token in Kleinbuchstaben um, entfernen gängige Stoppwörter und leiten Token zu ihrer Stammform ab.
Suchindex	Elasticsearch speichert alle Daten in einem Suchindex. Dieser Index wiederum ist aufgesplittet in Typen und Dokumente mit Eigenschaften.
SWAP-Modus / Swapping	In der Informatik beschreibt Swapping (engl. für Umlagerung; von englisch to swap, ‚austauschen') innerhalb der Speicherhierarchie das Auslagern von Daten einer höheren in eine tiefere Ebene oder das entgegengesetzte Laden. Durch Swapping sollen die Vorteile höherer Ebenen, also höhere Geschwindigkeit, und tieferer Ebenen, nämlich höhere Kapazität und kostengünstigerer Speicher, quasi gleichzeitig nutzbar gemacht werden.(Quelle:https://de.wikipedia.org/wiki/Swapping, Stand, 29.03.2022)
Token	Beim Erstellen oder Aktualisieren eines Datensatzes zerlegt die Elasticsearch den durchsuchbaren Text des Datensatzes in einzelne Begriffe, die sogenannten Token. Anschließend werden die Token im Suchindex gespeichert. Eine Suchanfrage wird wiederum in Token zerlegt. Nach diesen wird dann im Suchindex gesucht. Ein Token entspricht nicht immer einem grammatikalisch korrekten Wort, sondern ist eine Zerlegung und Vereinfachung von Worten, sodass auch eine Suche nach *Auto* ein Dokument findet, das nur das Wort *Autos* enthält.
Tomcat	Apache Tomcat ist ein Open-Source-Webserver und Webcontainer, der die Spezifikation für Java Servlets und Jakarta Server Pages (JSP) implementiert und es damit erlaubt, in Java geschriebene Web-Anwendungen auf Servlet- beziehungsweise JSP-Basis auszuführen.
YML	YML ist die Dateierweiterung der vereinfachten Auszeichnungssprache YAML. YAML wird zur Datenserialisierung verwendet und lehnt sich an XML und die Datenstrukturen in den Sprachen Perl, Python und C an.

Indexaufbau

Indexaufbau

Beim Indizieren eines Dokuments aus dem ELO Repository werden sowohl die Metadaten, als auch Volltextinhalte aus dem Dokument sowie Informationen aus Feed-Beiträgen in den Index der iSearch aufgenommen.

Dabei werden während des Indiziervorgangs Volltext- und Feed-Inhalte sprachabhängigen Analyseschritten unterworfen und das Ergebnis im jeweiligen Index der Elasticsearch gespeichert. Bei den Metadaten werden sowohl Analyseschritte bei Textfeldern durchgeführt als auch für jedes Feld die Inhalte unverändert (z. B. als Keyword, Datum usw.) gespeichert.

Dies erlaubt später, sehr flexibel für den jeweiligen Anwendungsfall nach den gewünschten Informationen zu suchen. Dabei kann gleichzeitig in Volltext- und Feed-Inhalten und Metadaten gesucht werden.

Vorbereitende Schritte zur Indizierung

Die Metadaten eines Dokuments aus dem ELO Repository liegen in der SQL-Datenbank vor und sind für den Indexserver sofort zugreifbar. Die Volltextinformationen dagegen müssen aus einem Dokument erst extrahiert werden. Dafür ist der ELO Textreader als eigenständiges Modul zuständig.

Der Textreader holt das jeweilige Dokument vom Indexserver, extrahiert daraus den Volltextinhalt und übergibt diesen wieder an den Indexserver, der ihn im ELO Repository als jeweils eigene Datei abspeichert. Der Textreader greift dabei, wenn nötig (z. B. bei Bilddateien) auf OCR-Software zurück, um daraus textuelle Inhalte zu extrahieren.

Regelmäßiger Indizierungsprozess

Der Indizierungsprozess wird durch den Indexserver durchgeführt. In regelmäßigem Zeitabstand (Updater-Intervall konfigurierbar über die ELO iSearch Configuration) prüft der Indexserver, welche Dokumente seit der letzten Indizierung hinzugekommen, geändert oder gelöscht worden sind. Diese müssen in den Indexen der iSearch aktualisiert werden.

Für die ermittelten Dokumente holt der Indexserver die Metadaten sowie Feed-Informationen aus der SQL-Datenbank. Den Inhalt des Volltextes (sofern vorhanden) holt der Indexserver aus dem Repository über den ELO Textreader. Die Informationen werden dann in aufbereiteter Form an die iSearch, d. h. an die Elasticsearch-Instanz übergeben, die dann die Feldinhalte indiziert. Dabei kommen die oben erwähnten sprachabhängigen Analyseschritte bei Textfeldern und Volltextinhalt zum Einsatz.

> **Beachten Sie**
>
> Wägen Sie gut ab, welche Dokumente in die iSearch aufgenommen werden sollen. Zu viele Textinformationen wirken sich negativ auf die Leistung der Such-Engine aus. Es kann deshalb sinnvoll sein, bestimmte Arten von Dokumenten auszuschließen (z. B. vollständige Bücher, Broschüren oder ähnliches).

Indizierung im Hintergrund

Einleitung

Dieses Dokument beschreibt die Funktion *Upgrade Index* des Serversetups, ihren Anwendungsfall und Bedienung.

> **Beachten Sie**
>
> Die Funktion *Upgrade Index* des ELO Server Setups darf ab ELO 12 nicht gegen die Produktivdatenbank durchgeführt werden, da im Laufe des Setups Änderungen an der Datenbank getätigt werden, mit denen ein älteres ELO aus dem Tritt geraten wird.
>
> Soll die Funktion dennoch genutzt werden, müssen die Daten der Produktivdatenbank in eine separate Datenbank-Instanz dupliziert werden. Beim *Upgrade Index* muss dann diese separate Datenbank-Instanz als Datenbank angegeben werden.
>
> Mit dieser Anleitung darf zudem nicht installiert werden. Der Indexaufbau für den Volltext findet vor der Installation statt und sollte vor Beginn der Installation abgeschlossen sein.

Funktionsbeschreibung

Bei der Aktualisierung der ELO iSearch auf eine neue Version, muss eine Re-Indexierung der Volltextdatenbank vorausgehen. Um die Verfügbarkeit der Volltextsuche währenddessen zu gewährleisten, erlaubt es die Funktion *Upgrade Index* im ELO Server Setup die Volltextdatenbank vor dem Update des Servers aufzubauen.

Hierzu wird ein abgeschottetes System, bestehend aus einem Apache Tomcat mit ELO Indexserver und einer ELO iSearch installiert, in welchem die Re-Indexierung stattfinden kann, während die ELO Serverkomponenten auf dem bisherigen Stand weiterlaufen. Nach Abschluss der Re-Indexierung kann der ELO Server mit wenig Ausfallzeit aktualisiert werden.

Anwendungsfall

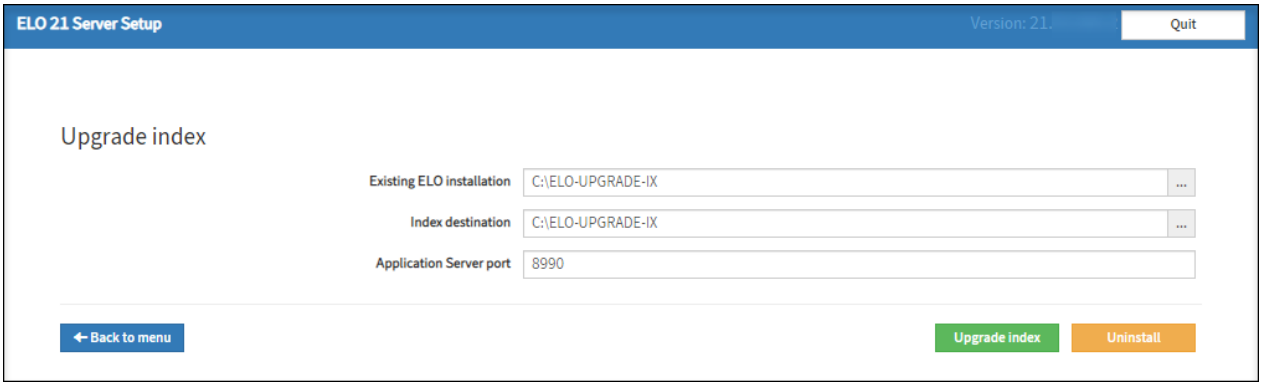
Die Funktion *Upgrade Index* sollte dann eingesetzt werden, wenn die Dauer der vollständigen Re-Indexierung und die damit einhergehende eingeschränkte Verfügbarkeit der Volltextsuche die Akzeptanzschwelle überschreiten würde.

Eine Re-Indexierung von bis zu ca. 1 Million Dokumenten (Summe aller Repositorys) sollte über ein Wochenende abgeschlossen sein. Dies ist selbstverständlich abhängig von der verwendeten Hardware und der durchschnittlichen Dokumentgröße.

Installation

Um die Re-Indexierung zu starten, starten Sie die aktuelle Version des ELO Server Setups.

Wählen Sie auf der Startseite des ELO Server Setups *Upgrade Index* aus.



Existing ELO installation: Pfad im Dateisystem zur vorhandenen ELO Installation.

Index destination: Pfad im Dateisystem, wohin das System zum Indexaufbau installiert werden soll. Zwecks der Übersichtlichkeit und des Aufräumens nach Abschluss der Re-Indexierung empfiehlt es sich, hier einen von der ELO Installation unterschiedlichen Pfad anzugeben.

Application Server Port: Port für den Tomcat des Indexservers. Dieser darf nicht schon belegt sein durch einen anderen Dienst des Betriebssystems. Dies schließt ELO Dienste mit ein.

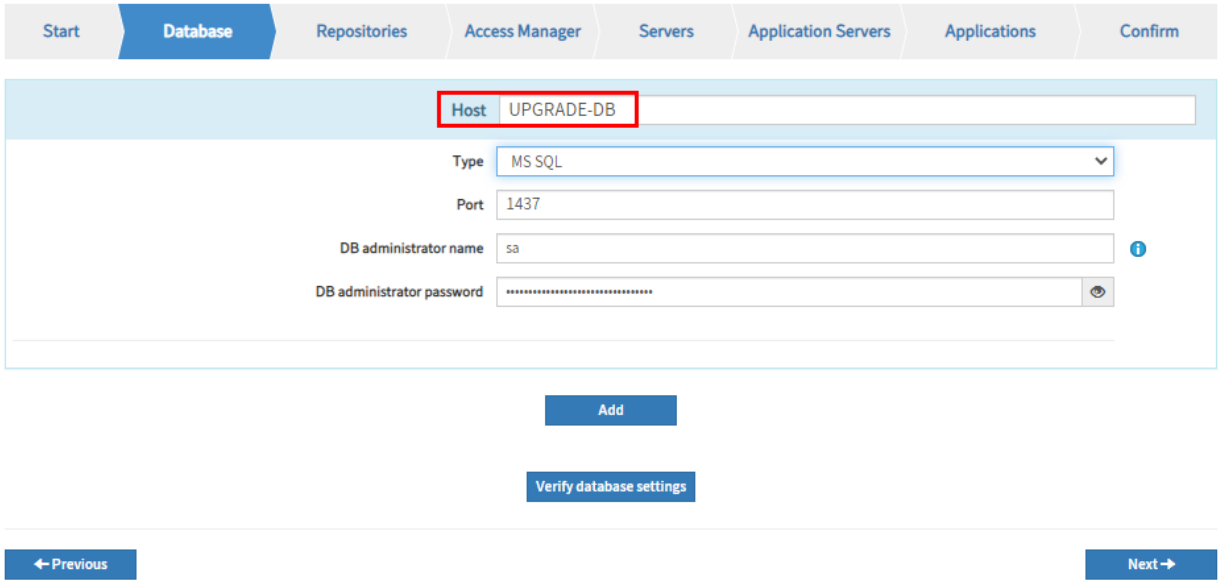
Upgrade Index: Konfiguriert das System für die Re-Indexierung, es wird zunächst auf eine Zusammenfassung weitergeleitet, von dort kann die eigentliche Installation dann gestartet werden.

Uninstall: Deinstalliert das bei *Upgrade Index* installierte System. Dies ist nach Abschluss der Re-Indexierung durchzuführen. Es werden die Dienste und Programme entfernt, die für die Re-Indexierung installiert worden sind. Währenddessen produzierte Daten wie die Volltextdatenbank bleiben erhalten und müssen im Anschluss manuell entfernt werden.

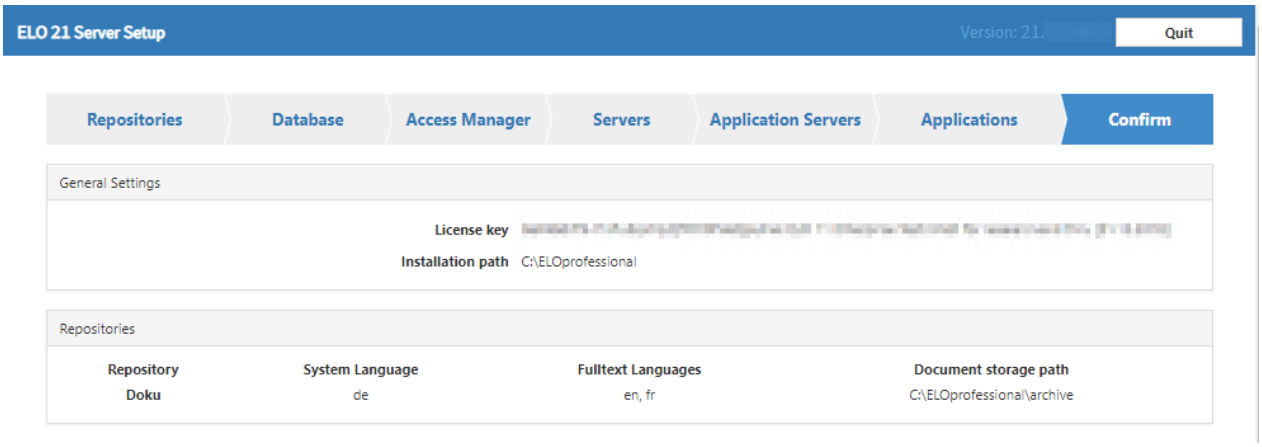
Upgrade Index

Durch Auswählen des Buttons *Upgrade Index* werden Sie auf die Übersichtsseite der Installation weitergeleitet. Die dort zusammengefassten Angaben betreffen nur an relevanten Stellen den Vorgang der Re-Indexierung.

> **Beachten Sie**
>
> Die Funktion *Upgrade Index* des ELO Server Setups darf ab ELO 12 nicht gegen die Produktivdatenbank durchgeführt werden, da im Laufe des Setups Änderungen an der Datenbank vorgenommen werden, mit denen ein älteres ELO nicht mehr fehlerfrei betrieben werden kann (hat z. B. Auswirkungen auf die Vertretungsregeln).



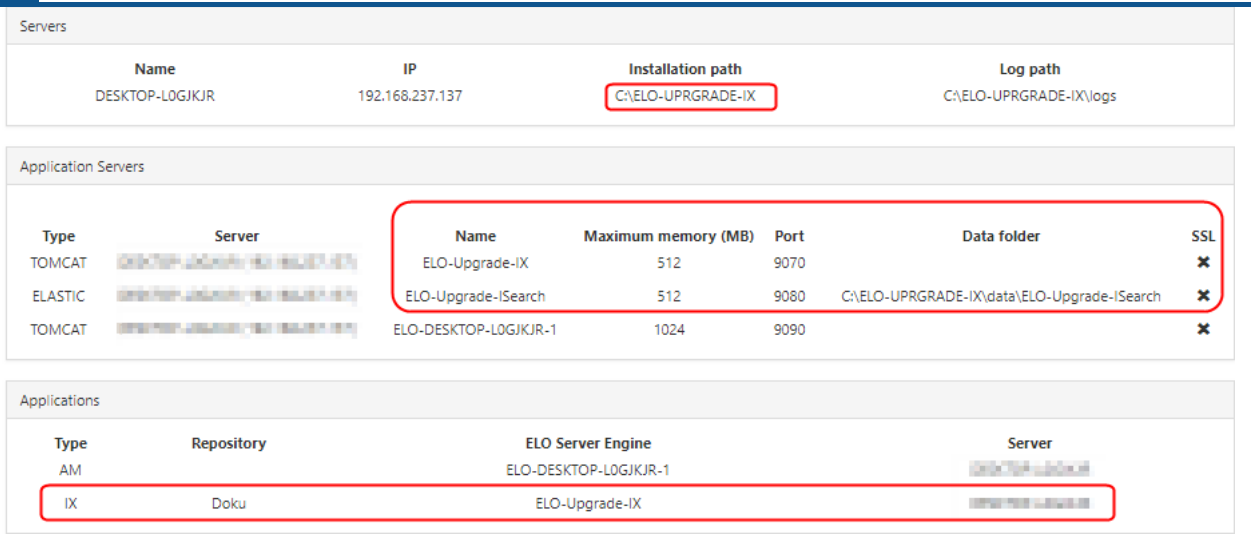
Möchten Sie die Funktion dennoch verwenden, müssen Sie die Daten der Produktivdatenbank in eine separate Datenbank-Instanz duplizieren. Beim Upgrade Index müssen Sie dann im Tab *Database* im Feld *Host* diese separate Datenbank-Instanz angeben.



Der hier angegebene Installationspfad und die Dokumentenpfade müssen den Pfaden der ELO Installation entsprechen.

Beachten Sie

Sie müssen einen aktuellen Lizenzschlüssel verwenden.



Die Pfade (*Installation path*) unter *Servers* müssen dem Pfad entsprechen, der zuvor bei *Index destination* festgelegt wurde. Dies gilt ebenso für den *Data Folder* unter *Application Servers*. An dieser Stelle wird nach der Re-Indexierung die Volltextdatenbank der ELO iSearch zu finden sein.

Wählen Sie *Install*, um die Installation des Upgrade-Systems zu starten.

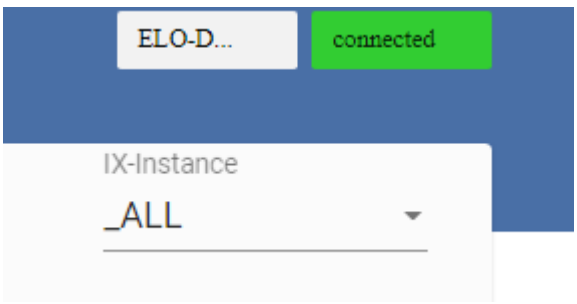
Das ELO Server Setup installiert das Upgrade-System bestehend aus einem ELO Indexserver und einer Instanz der ELO iSearch.

> **Information**
>
> Dabei werden keine Änderungen am bisherigen System durchgeführt.

Nach dem Abschluss der Installation schließen Sie das ELO Server Setup mit dem Button *Finish*.

Re-Indexierung im Upgrade-System

Ist die Installation des Upgrade-Systems abgeschlossen, rufen Sie die Seite *ELO Elasticsearch Configuration* auf (über die *Statusseite des neu installierten ELO Indexservers > Configure Options > iSearch config*).

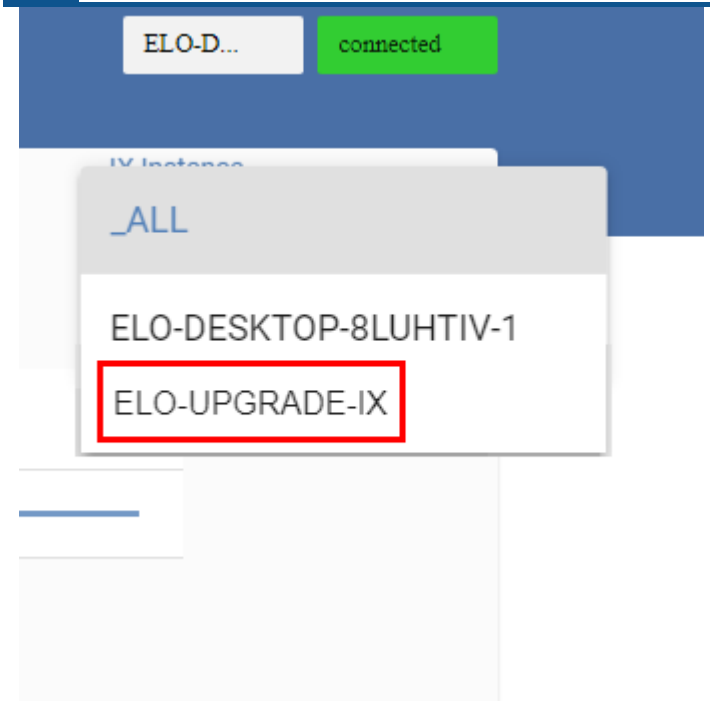


Wählen Sie im Drop-down-Menü *IX-Instance* die Instanz *_ALL* aus, falls noch nicht ausgewählt.

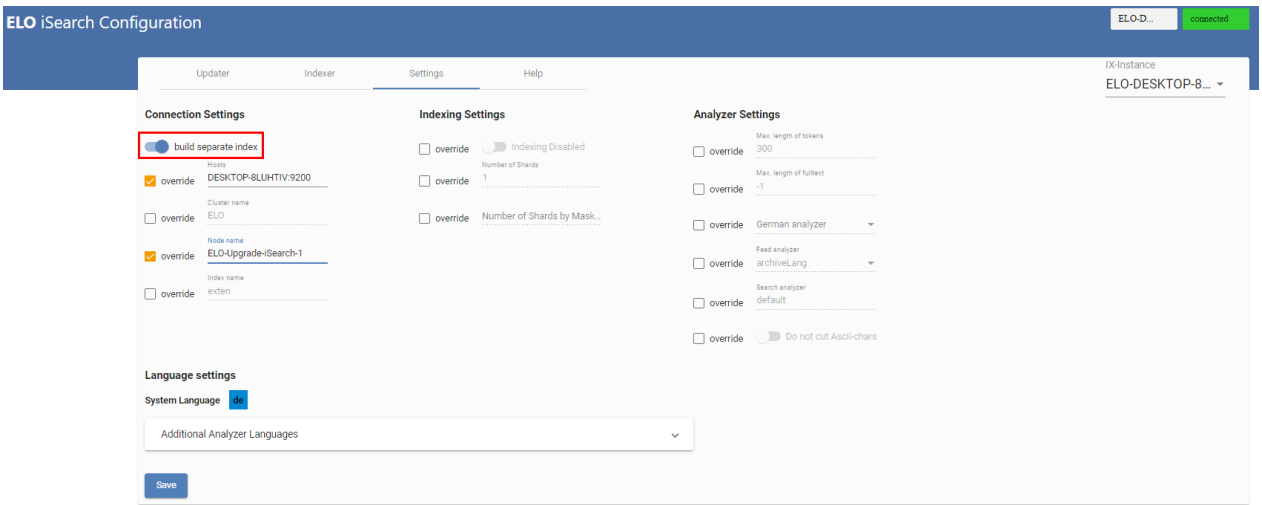
Prüfen Sie auf dem Tab *Settings* die Einstellungen für die Instanz *_ALL*.

Achten Sie besonders darauf, ob im Bereich *Language settings* die richtigen Sprachen ausgewählt sind.

Speichern Sie die Einstellungen mit *Save*.



Wählen Sie über das Drop-down-Menü *IX-Instance* die Instanz des Upgrade-Systems aus.



Prüfen Sie die Einstellungen für die Instanz auf dem Tab *Settings*.

Im Feld *Index name* muss der Name des Repositorys (in Kleinbuchstaben) eingetragen sein.

Aktivieren Sie den Schalter *build seperate index*.

Speichern Sie die Einstellungen mit *Save*.

Prüfen Sie die Einstellungen auf dem Tab *Indexer*

Die Felder *From object id* und *To object id* müssen mit dem Wert `-1` ausgefüllt sein.

Speichern Sie die Einstellungen mit *Save*.

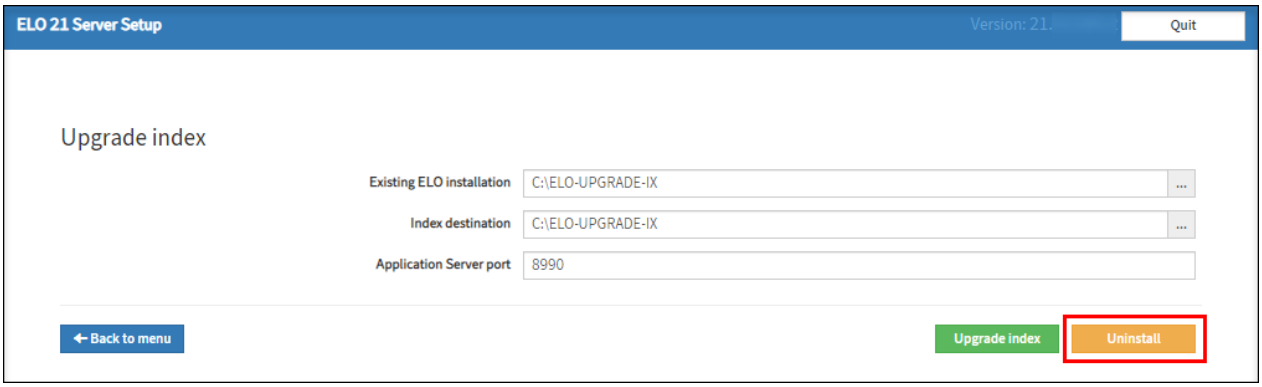
Starten Sie die Re-Indexierung mit dem Button *Create new index*.

Deinstallation des Upgrade-Systems

Nachdem der Re-Indexierungsprozess abgeschlossen ist, können Sie das Upgrade-System deinstallieren. Der erstellte Index bleibt dabei erhalten.

Starten Sie dazu erneut das ELO Server Setup.

Wählen Sie auf der Startseite des ELO Server Setups *Upgrade Index*.



Die Seite Upgrade Index erscheint. Die Pfadangaben sollten dem zuvor erstellten Upgrade-System entsprechen.

Wählen Sie *Uninstall* aus.

Das Upgrade-System wird deinstalliert.

Nach dem Abschluss der Deinstallation schließen Sie das ELO Server Setup mit dem Button *Finish*.

Update auf neue ELO Version

Ist das Upgrade-System erfolgreich deinstalliert, können Sie nun den eigentlichen ELO Server updaten.

Führen Sie dazu erneut das ELO Server Setup aus.

Aktualisieren Sie den ELO Server wie gewohnt (nähere Informationen finden Sie in der Dokumentation *ELO Server*) über die Funktionen *Standard Configuration/Update* oder *Advanced Configuration/Update*.

Nach dem Abschluss der Installation schließen Sie das ELO Server Setup mit dem Button *Finish*.

Migration der Daten

Haben Sie die neue ELO Version erfolgreich installiert, können Sie den vom Upgrade-System erstellten Index ins Produktivsystem übernehmen.

Stoppen Sie den Dienst für die ELO iSearch.

Navigieren Sie im Dateisystem zu folgendem Ordner:

`<Installationspfad des Upgrade-Systems>/data/ELO-Upgrade-ISearch/`

Kopieren Sie den Order *index*.

Navigieren Sie im Dateisystem zu folgendem Ordner:

`<Installationspfad des ELO Servers>/data/`

Fügen Sie den kopierten Ordner ein und ersetzen Sie den vorhandenen Ordner *index*.

Starten Sie den Dienst der ELO iSearch.

Prüfen Sie die Einstellungen auf der Seite *ELO Elasticsearch Configuration.*

Achten Sie darauf, dass für keine Instanz der Schalter *build separate index* aktiviert ist.

Testen Sie die ELO iSearch über einen ELO Client.

War die Migration erfolgreich, können Sie die verbliebenen Daten des Upgrade-Systems löschen.

Upgrade Index auf getrenntem Server

Die Re-Indexierung benötigt sehr viele CPU-Ressourcen und kann einen Server stark auslasten. Zum einen kann durch Reduktion der Prozesspriorität der Einfluss auf das Produktivsystem reduziert werden, zum anderen durch Ausführen der Re-Indexierung auf einem getrennten Server.

Das Setup benötigt beide Konfigurationsdateien (`%HOMEPATH%\.elosetup.conf` sowie `<ELO>\config\serversetup2\elosetup.conf`), um die Funktion *Upgrade Index* auf einem anderen Server auszuführen. Der Ordner `<ELO>\config\serversetup2\` ist dabei anzulegen, der Pfad muss dem des produktiven Systems gleichen. Weitere Ordner wie z. B. `<ELO>\archive` oder `<ELO>\data` werden nicht benötigt.

Im Anschluss kann die Funktion wie in Upgrade Index beschrieben ausgeführt werden.

Upgrade Index bei Migration von ELO 9

Die Volltextdatenbank kann auch bei einem Wechsel von ELO 9 auf ELO 10.1 oder neuer mit diesem Verfahren vorbereitet werden. Durch das Ausrollen des Indexservers 10 wird auch die Datenbank aktualisiert. Daher ist bei ELO 9 mindestens Version 9.16.110 des Indexservers Voraussetzung. Ein Update von ELO 8 auf ELO 10.1 wird nicht unterstützt.

Wie in Upgrade Index auf getrenntem Server beschrieben, benötigt das Setup für die Funktion *Upgrade Index* zwei Konfigurationsdateien, die es bei ELO 9 Systemen noch nicht gibt. Diese müssen zuvor erzeugt werden. Das Setup legt die beiden Dateien bei einem Upgrade an.

> **Achtung**
>
> Das Upgrade darf nicht vollständig durchgeführt werden. Nach der Erfassung des existierenden Installationspfades sowie des Lizenzschlüssels müssen Sie nach der Analyse des Systems mit der Auswahl von *Configure Upgrade* oder *Review Upgrade* die Upgrade-Prozedur abbrechen. Es darf nicht installiert werden!

Im Anschluss sind die benötigten Dateien vorhanden. Nun kann mit Punkt Upgrade Index oder Upgrade Index auf getrenntem Server fortgefahren werden.