

Erweiterung der TextGrid-Autorisierungsinfrastruktur für die Unterstützung von Lizenzen

Version 3 (30.6.2010)

Oliver Schonefeld, Andreas Witt, Markus Widmer, Peter Gietz

Einleitung

In TextGrid gibt es verschiedene Content-Provider, deren Ressourcen nicht ohne weiteres in der TextGrid-Infrastruktur zur Verfügung gestellt werden können. Die Ursache hierfür ist, dass die erforderlichen Zugriffsbeschränkungen bislang nicht von der existierenden Autorisierungsinfrastruktur abgebildet werden können. Beispielsweise ist es für den Zugriff auf einige Ressourcen am Institut für Deutsche Sprache notwendig, dass Benutzer einen Lizenzvertrag akzeptieren. Um diesen Content-Providern die Bereitstellung ihrer Ressourcen in TextGrid zu ermöglichen, muss die bestehende Autorisierungsinfrastruktur erweitert werden, um feinere Zugriffsbeschränkungen zu ermöglichen.

Beispielszenarien

Verschiedene Anwendungsfälle sollen die Anforderungen an das System skizzieren.

- Ein Benutzer muss einer bestimmten Organisation angehören, um auf eine Ressource zugreifen zu können.
- Ein Benutzer muss sich an einem bestimmten Ort aufhalten, z.B. dem Forschungsinstitut oder einem Land, um auf eine Ressource zugreifen zu können. (Hierunter fällt auch das umgekehrte Fall: das ‚Embargoszenario‘)
- Ein Benutzer muss erst einen Lizenzvertrag bestätigen um auf eine Ressource zugreifen zu können.
- Ein Benutzer darf erst nach einem bestimmten Datum auf eine Ressource zugreifen, z.B. 75 Jahre nach dem Tod des Autors.
- Ein Benutzer darf erst nach einer bestimmten Zeitspanne (z.B. 6 Monate) nachdem die Ressource erzeugt wurde, auf diese Ressource zugreifen.
- Eine beliebige Kombination der oben genannten Bedingungen

Modellierung

Lizenzverträge (bzw. andere komplette Zugriffsbeschränkungen) werden in einem Modell abgebildet, das diesen Vertrag in einem Satz von Regeln modelliert, die auf die Eigenschaften, d.h. Attribute, der Ressource und des Benutzers angewendet werden. Eine Ressource kann potentiell mehreren Lizenzverträgen zugeordnet sein. In diesem Fall wird der Zugriff auf die Ressource dann gewährt, wenn der Benutzer die Bedingungen von mindestens einem Vertrag erfüllt. Etwas formeller ließe sich die Modellierung wie folgt beschreiben:

- eine Ressource hat eine Reihe von Attributen, z.B. Erstellungsdatum oder Besitzer. Diese ergeben sich zum größten Teil aus den (TextGrid-)Metadaten. Ggf. muss noch geprüft werden, ob Content-Provider eigene Attribute definieren könnten. Aber auch diese ließen sich in den (TextGrid-)Metadaten speichern.
- ein Benutzer hat eine Reihe von Attributen, wie z.B. Institutszugehörigkeit. Diese Attribute ergeben sich einerseits aus den Shibboleth-Attributen, die der IDP bei der Authentisierung übermittelt und andererseits aus anderen Informationen, z.B. IP-Adresse des Benutzers. Ggf. müssen weitere Attribute zu einem Benutzer auch noch innerhalb der TextGrid-Autorisierungsinfrastruktur gespeichert werden (siehe nächster Abschnitt).
- ein Lizenzvertrag ist ein Satz von Regeln, die die Attribute von Ressource und Benutzer in Beziehung setzen. Diese Regeln werden beim Zugriff eines Benutzers auf die Ressource ausgewertet. Ein Lizenzvertrag kann mehreren Ressourcen und eine Ressource kann mehrere Lizenzverträgen zugeordnet sein.
- Die Lizenzverträge werden dezentral als XML-Dokument (gegebenenfalls XACML) verwaltet und in eine zentrale Datenbank synchronisiert.

Die Autorisierungsinfrastruktur, d.h. die Auswertungsmaschinerie, sollte prinzipiell mit einer beliebigen Anzahl von Regeln in Lizenzverträgen und Attributen bei Ressourcen und Benutzern zurechtkommen, da zum aktuellen Zeitpunkt nicht genau definiert werden kann, welche Regeln und Attribute benötigt werden. Es ist jedoch durchaus sinnvoll einen Satz von Relationen in den Regeln (gleich, größer, kleiner, ...) und Datentypen (Zeichenkette, Datum, Boolean, ...) zu definieren.

Probleme, Anmerkungen und Implementierungsideen

Bei Lizenzverträgen ist es meist so, dass Benutzer diesen akzeptieren müssen, d.h. es muss irgendwo hinterlegt werden, ob ein Benutzer einen bestimmten Vertrag akzeptiert hat oder nicht. Prinzipiell ließe sich dies in Shibboleth-Attributen (wie z.B. eduPersonEntitlement) abbilden, jedoch ist es wahrscheinlich eher problematisch, besonders großen Organisationen wie Universitäten, davon zu überzeugen die Benutzerdaten regelmäßig und zeitnah um TextGrid-spezifische Attribut(-Werte) zu aktualisieren. Daher ist es ggf. sinnvoll (und schneller zielführend) solche Attribute innerhalb der TextGrid-Infrastruktur selbst zu speichern. Oft ist es auch so, dass das Akzeptieren eines Lizenzvertrags bei einem Ressourcenbereiter „offline“ erfolgt, z.B. durch Zusendung eines unterschriebenen Lizenzvertrags per Telefax und entsprechende Nachweise dort gespeichert werden.

Daher ist es sinnvoll, dass Ressourcenbereiter sowohl Lizenzverträge als auch relevante Informationen zu Benutzern (wie z.B. ob ein bestimmter Lizenzvertrag akzeptiert wurde) dezentral vorhalten. Um eine schnelle Entscheidung für eine Autorisierungsanfrage innerhalb von TextGrid beantworten zu können, müssen diese Daten zeitnah über einen Synchronisierungsmechanismus an die Autorisierungskomponente übermittelt werden. Der Mechanismus muss sowohl das Erstellen neuer Daten, als auch das Ändern und Löschen bestehender Daten ermöglichen. Es ist noch zu prüfen, ob es sinnvoller ist, diesen Mechanismus besser als Push- oder Pull-Verfahren auszulegen und ob Lizenzverträge und Benutzerinformationen über einen gemeinsamen oder einen trennten Kanal synchronisiert werden.

Um eine bessere Performance von TG-Auth zu gewährleisten, werden die Shibboleth-Attribute als auch weitere relevante Informationen zu einem Benutzer (z.B. die IP-Adresse) beim Erstellen einer Session in TG-Auth hinterlegt. Weiterhin können natürlich nur alle Ressourcen in Lizenzverträge aufgenommen werden, die in TG-Auth bekannt sind. Für diese kann dann an zentraler Stelle eine Entscheidung durch den PDP getroffen werden können.

Örtliche Zugriffsbeschränkungen ließen sich in erster Näherung durch Abgleich der IP-Adresse des Benutzers mit vordefinierten Listen (im Falle von Organisationen) oder der GeoIP-Datenbank (im Falle von Ländern) implementieren. VPN-Verbindungen oder ähnliches würden von einer solchen Implementierung jedoch nur bedingt abgedeckt.

Die rechtlichen Konsequenzen, die sich durch die Modellierung und die skizzierte Implementierung z.B. hinsichtlich Haftungsregelungen und ggf. notwendiger Verträge ergeben, müssen noch geprüft werden.