

# Inhalte schneller finden: Die Terminologie der Technischen Dokumentation steigert den Nutzen von Microsoft SharePoint

## White Paper

### Motivation:

Das Wissensmanagement mit Microsoft SharePoint leidet wenn die Suche nach Inhalten länger als nötig dauert oder nur suboptimale Suchergebnisse liefert. Die Qualität und Effizienz der Suchanwendung und damit des Wissensmanagements kann verbessert werden, wenn (1) die vorhandenen Inhalte adäquater erschlossen wären oder (2) stärkere Suchwerkzeuge eingesetzt werden könnten. Beide Punkte werden durch den Einsatz von kontrollierten Terminologie-Sammlungen adressiert. Denn ein kontrolliertes Vokabular bzw. gesicherte Terminologie-Sammlungen sind in SharePoint notwendig, um Inhalte effizient zu erschließen, die Qualität der Metadaten zu steigern bzw. effektive Indexierung und Verschlagwortung zu ermöglichen. In vielen Unternehmen stehen bereits Terminologie-Sammlungen zur Verfügung, die als Basis für die Verschlagwortung in SharePoint wiederverwendet werden können. Zum Beispiel kann die Technische Dokumentation als optimaler Ausgangspunkt für den Aufbau einer initialen terminologischen Basis dienen, die u. a. in SharePoint verwendet werden kann.

Mit diesem White Paper möchten wir Entscheidungsträger und Praktiker auf unerschlossene Quellen von terminologischen Daten im Unternehmen hinweisen, eine Methode zur Integration dieser Wissensquelle in die SharePoint-Informationsarchitektur aufzeigen und ihren erheblichen Beitrag zum Nutzen von SharePoint konkretisieren.

### Autoren:

Dott.ssa Micaela Andrich, RWS Group Deutschland GmbH

Dr. Michael Erdmann und Daniel Hansch, DIQA Projektmanagement GmbH



© RWS Group GmbH und DIQA Projektmanagement GmbH 2014

Kontakt: DIQA Projektmanagement GmbH

Pfinztalstraße 90

76227 Karlsruhe

<http://www.diga-pm.de>

## Fehlende terminologische Ordnung begrenzt den effizienten Einsatz von SharePoint

Historisch eine webbasierte Content- und Dokumentmanagement-Plattform, ist Microsoft SharePoint heute eine kollaborative und soziale Anwendung, die die Aktivitäten des Wissensmanagements auf vielfältige Weise unterstützt und als Drehscheibe der Geschäftsprozesse dient.

### Welche Rolle spielen Metadaten und Terminologie-Sammlungen in SharePoint?

Mit Hilfe von technischen und inhaltlichen Metadaten werden in SharePoint Dokumente, Blog- und Wiki-Einträge, sowie Listenelemente und externe Inhalte erschlossen. Häufig sind die vorhandenen Metadaten unvollständig, inkonsistent oder sogar falsch. Der wichtigste Grund für die mangelhaften (inhaltlichen) Metadaten liegt darin begründet, dass sie von Benutzern oft händisch erzeugt werden müssen ohne jegliche Unterstützung von vorhandenen Strukturen und Kontrollmechanismen (z. B. *Spellchecker*). Das führt einerseits zu inkonsistenten Metadaten aufgrund subjektiver Entscheidungen von Einzelnen, andererseits werden Metadaten nur in sehr geringem Umfang erzeugt („ökonomisches“ Nutzerverhalten). Wird die Eingabe von Metadatenfeldern durch das System erzwungen, kommt es vor, dass Benutzer beliebige Werte eintragen oder den ersten Eintrag aus der Auswahlliste wählen, so dass ungenaue, oder sogar irreführende Metadaten entstehen.

Unvollständige, inkonsistente und falsche Metadaten führen daher dazu, dass Inhalte nicht adäquat erschlossen werden. Sie verhindern dann, dass die gesuchte Information effizient gefunden wird, denn sie wirken sich negativ auf die sogenannten Such-*Refiner* für die facetiierte Navigation aus. Metadaten als Filter verlieren ihren Nutzen, wenn aufgrund synonyme oder fehlerhafter Schreibweisen beispielsweise ein Produkt mehrfach als Einschränkungsmöglichkeit auftaucht. Das Risiko inkonsistenter Metadaten lässt sich durch die Verwendung von kontrollierten Terminologie-Sammlungen mindern. Sie sind auch die Voraussetzung für den Einsatz der ausgefeilten Suchmethoden, die SharePoint zur Verfügung stellt, wie beispielsweise der *Advanced Search*, der Suche mit Synonymen und Unter- und Oberbegriffen und der unternehmensweiten Suche.

### Hindernisse für die Nutzung von terminologischen Daten in SharePoint

Leider ist die Verwendung von Terminologie-Sammlungen in SharePoint keine Selbstverständlichkeit, denn das Terminologie-Management in SharePoint ist umständlich und zeitraubend. Beispielsweise werden Änderungen an der Terminologie in SharePoint sofort bei den verwalteten Metadaten wirksam, d.h. größere Überarbeitungen existierender Terminologie-Bestände sollten außerhalb von SharePoint durchgeführt werden. In der Praxis weicht man daher oft auf Excel oder Texteditoren aus, obwohl sie keine adäquaten Terminologie-Management-Features bieten. Da SharePoint keines der

gängigen Formate für den Import von Terminologie-Sammlungen (z. B. SKOS, TBX) unterstützt, wird auch die Wiederverwendung existierender Terminologie-Sammlungen erschwert.

Offt wird auch aus organisatorischen Gründen der Mehrwert von Terminologie-Sammlungen nicht erkannt, weil die Rolle des Informationsarchitekten nicht adäquat besetzt ist, sondern nebenher von einem technischen Administrator wahrgenommen wird. Die fehlende Terminologie-Expertise im SharePoint-Team führt auch dazu, dass Terminologie-Sammlungen nicht selbst aufgebaut werden können. Mangelnder Wissenstransfer zwischen Abteilungen verhindert, dass existierende Terminologie-Listen und Taxonomien, wie z. B. aus der Technischen Dokumentation, als wichtige Quelle identifiziert und genutzt werden.

## **Die Technische Dokumentation als Quelle für kontrollierte Terminologie-Listen und Taxonomien**

### **Terminologie-Listen der Technische Redaktion**

Jede in Europa verkaufte Maschinen und jedes technische Produkt muss von einer technischen Dokumentation begleitet werden. Dazu gehören Betriebs-, Montage-, Wartungs- und Reparaturanleitungen, die wiederum in diverse Sprachen übersetzt werden müssen. Als Texttypologie gehören diese Schriften, zusammen mit den Dokumenten, die im Produktzyklus entstehen (z. B. Risikobeurteilungen, Versuchsberichte, technische Zeichnungen), zum expliziten Wissen des Unternehmens. In modular aufgebauten Inhaltsverzeichnissen externalisieren diese Dokumente den technischen „Status quo“ des Unternehmens bzw. den Wissensvorsprung des Unternehmens gegenüber Konkurrenzunternehmen und fassen die relevante unternehmensspezifische Terminologie, u. a. Produktnamen, Produktgruppen, Komponenten, Verfahren (s. Tabelle 1), zusammen, die wiederum in anderen Systemen vorkommt (u. a. PIM-, PLM- oder ERP-Systeme).

### **Terminologie-Listen für den Übersetzungsprozess**

Als Gesamtheit der Fachbegriffe (Termini) eines Fachgebietes ist die Terminologie die Basis jeglicher Fachkommunikation bzw. ein Mittel für den effizienten und effektiven Wissenstransfer unter den Fachexperten. Die systematische Erfassung, Verwaltung und Verbreitung korrekter und konsistenter Fachterminologien wird Terminologie-Management (TM) genannt. Im Unternehmen wird diese Tätigkeit von den technischen Redakteuren bzw. Terminologen übernommen, die Terminologie-Listen in Redaktions- bzw. Terminologie-Datenbank-Systemen verwalten, die wiederum in Übersetzungs- und Autorenwerkzeugumgebungen eingebunden sind. Ein direkter belegbarer Mehrwert der Pflege von Terminologie lässt sich durch die Reduktion von Übersetzungskosten und Produkteinführungszeit (*Time-to-Market*) dank kuratierter Terminologie-Sammlungen beweisen. Die Mehrwerte der Verwaltung von Terminologie lassen sich aber auch in anderen Unternehmensbereichen spüren, wie Vertrieb, Marketing, Produktmarketing, Customer Support, Training, Human Resources und nicht zuletzt im Innovations- und Wissensmanagement.

<b>Begriff</b>	Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird (DIN 2342 1/10.92)	<b>Heizkreisverteiler</b> [Fachgebiet: Wärme- und Kältetechnik]
<b>Benennung</b>	aus einem Wort oder mehreren Wörtern bestehende Bezeichnung (DIN 2342 1/10.92)	<b>Heizkreisverteiler</b>
<b>Hyperonym</b> skos:broader TBX: broaderconcept	Übergeordneter Begriff innerhalb eines hierarchischen Begriffssystems, das durch Abstraktionsbeziehungen gekennzeichnet ist. (DIN 2330/12.93)	<b>Verteiler</b> Heizkreisverteiler
<b>Hyponym</b> skos:narrower TBX: subordinateconcept	untergeordneter Begriff innerhalb eines hierarchischen Begriffssystems, das durch Abstraktionsbeziehungen gekennzeichnet ist. (DIN 2330/12.93)	Heizkreisverteiler <b>Vorlaufkammer, Rücklaufkammer, Heizkammer, Kesselkreis</b>
<b>Synonym</b> skos:prefLabel TBX: preferredterm	Wort von gleicher oder ähnlicher Bedeutung (Wahrig)	Allstrommotor Universalmotor [Fachgebiet: Kraftfahrzeugtec hnik]
<b>Terminus</b>	zusammengehöriges Paar aus einem Begriff und seiner Benennung als Element einer Terminologie (DIN 2330/12.93)	<b>Heizkreisverteiler</b>

Tabelle 1. Beispiele für terminologisches Wissen (mit Definitionen)

## Import kuratierter Terminologie-Sammlungen in SharePoint

Mit SharePoint (*out-of-the-box*) ist es allerdings kaum möglich, Terminologie-Sammlungen zu verwenden, die nicht in SharePoint selbst erstellt wurden. Um die reichen Terminologie-Sammlungen der technischen Redaktion einsetzen zu können, wird dementsprechend ein zusätzliches Werkzeug benötigt. Mit GRASP<sup>1</sup> steht ein solches zur Verfügung. Im Kontext dieses Artikels sind GRASPs Terminologie-Management-Funktionalitäten besonders relevant. SharePoint-Nutzer können damit Terminologie-Sammlungen in Form von TBX-Dateien importieren und so im TermStore neue TermSets anlegen.

### Was ist TBX?

TermBase eXchange (TBX)<sup>2</sup> ist ein XML-basiertes und von der ISO standardisiertes Austauschformat für die Repräsentation und Verbreitung von Terminologie-Daten (ISO 30042), das von den gängigen TM-Systemen unterstützt wird. Eine TBX-Datei ist begriffsorientiert. Sie enthält eine Sammlung

<sup>1</sup> <http://www.diaq-pm.de/de/GRASP>

<sup>2</sup> [http://www.gala-global.org/oscarStandards/tbx/tbx\\_oscar.pdf](http://www.gala-global.org/oscarStandards/tbx/tbx_oscar.pdf)

von einzelnen Einträgen, die jeweils einen Begriff darstellen, samt Benennungen und Metadaten ggf. in verschiedenen Sprachen. Im ISO 30042 stehen insgesamt 127 Datenkategorien zur Verfügung, die die Terminologiearbeit unterstützen sollen (z. B. Definition-Feld, Kontext-Feld, Vorzugsbenennung-Attribut). Von besonderer Bedeutung für die spätere Darstellung von Begriffssystemen bzw. das Ableiten von Taxonomien sind die Datenkategorien, die sich auf Beziehungen zwischen Begriffen oder Benennungen beziehen, u. a. *subordinateConcept*, *broaderConcept* und *relatedConcept*, die auch im SKOS-Format<sup>3</sup> vorkommen (*skos:narrower*, *skos:broader*, *skos:related*).

## TBX-Import

Der Import-Vorgang nach SharePoint übersetzt alle Termini und Benennungen, sowie andere Attribute des TBX-Modells in entsprechende TermStore-Einträge, so dass ein vollständiges TermSet inkl. hierarchischer Beziehungen, Synonyme und zielsprachigen Benennungen entsteht, das unmittelbar verwendet werden kann. Dabei wird SKOS als internes Zwischenformat verwendet, da GRASP in der Lage ist, beliebige SKOS-Formate direkt in den TermStore zu importieren.

Der Terminologie-Lebenszyklus geht natürlich außerhalb von SharePoint weiter, so dass es wichtig ist auch Updates der Terminologie einspielen zu können. GRASP sorgt beim Update-Prozess dafür, dass die Identität der Termini erhalten bleibt, damit bereits verwendete Termini weiterhin ihre Gültigkeit behalten und annotierte Dokumente weiterhin korrekt annotiert bleiben. Damit hilft GRASP diverse organisatorische Schwierigkeiten bei der (Wieder-)Verwendung von Terminologie-Sammlungen zu vermeiden:

- Die Terminologie-Sammlungen werden von Mitarbeitern der technischen Dokumentation kuratiert, die Expertise im Terminologie-Management besitzen;
- Ein technischer Mitarbeiter kann für den Import (einer neuen Version) sorgen, so dass inhaltliche und administrative Kompetenzen getrennt sind;
- Vor der Verwendung einer neuen Version kann sie auch in SharePoint visualisiert werden, so dass ein Update- und Abnahmeprozess in SharePoint möglich wird.

## Terminologie-Sammlungen steigern den Nutzen von SharePoint

### Governance

In SharePoint (z. B. in *Lookup*-Feldern) wird terminologisches Wissen in den meisten Fällen *ad-hoc* in Form von Auswahllisten verwaltet, die keiner unternehmensweiten *Governance* unterworfen sind. Auf diese Weise kann nicht sichergestellt werden, dass eine gemeinsame und konsistente Terminologie verwendet wird. Darüber hinaus unterstützt dieser Ansatz weder

---

<sup>3</sup> <http://www.w3.org/TR/skos-reference/>

die Verwaltung von Synonymie noch Term-Hierarchien. Die Verwendung von konsolidierten Terminologie-Sammlungen ermöglicht es, *Lookup*-Felder durch verwaltete Metadaten zu ersetzen, deren Termini (inkl. Synonymen und mehrsprachige Benennungen) zentral im TermStore gepflegt werden. Damit ist gewährleistet, dass inkonsistente Benennungen zuverlässig erkannt bzw. schnell korrigiert werden können.

Mit Hilfe von GRASP können Terminologie-Sammlungen direkt in SharePoint überarbeitet werden, ohne dass Benutzer auf ein externes, ungeeignetes Werkzeug wie Excel oder den Texteditor ausweichen müssen. Dazu erzeugt man in GRASP eine Arbeitskopie einer Terminologie-Sammlung, die die Fachbenutzer gemeinsam in SharePoint überarbeiten können. Die Produktivterminologie wird dann durch die Arbeitskopie ersetzt wenn sie fertig entwickelt und abgenommen ist. Erst dann wirken sich die geänderten Termini auf die verwalteten Metadaten aus.

## Autotagging

Terminologie-Sammlungen sind in SharePoint eine unerlässliche Wissensquelle, um Inhalte kostengünstig zu erschließen und die Qualität und Quantität der Metadaten zu steigern, denn sie ermöglichen die automatische Verschlagwortung von Dokumenten (z. B. mit dem AutoTagger<sup>4</sup> von Layer2). Damit werden inkonsistente Metadaten aufgrund subjektiver Entscheidungen Einzelner vermieden und die Benutzer entlastet.

## Suchverbesserungen

Wenn Dokumente und Listeneinträge gut durch Metadaten erschlossen sind, verbessern sich die Such-Ergebnisse automatisch, z. B. dadurch, dass in den Such-*Refinern* die einzelnen Facetten gut gefüllt sind, so dass facettierte Navigation die Suchtreffer systematisch einschränken kann (s. Abb. 1).

---

<sup>4</sup> <http://www.layer2.de/de/produkte/Seiten/SharePoint-2010-Auto-Tagger.aspx>

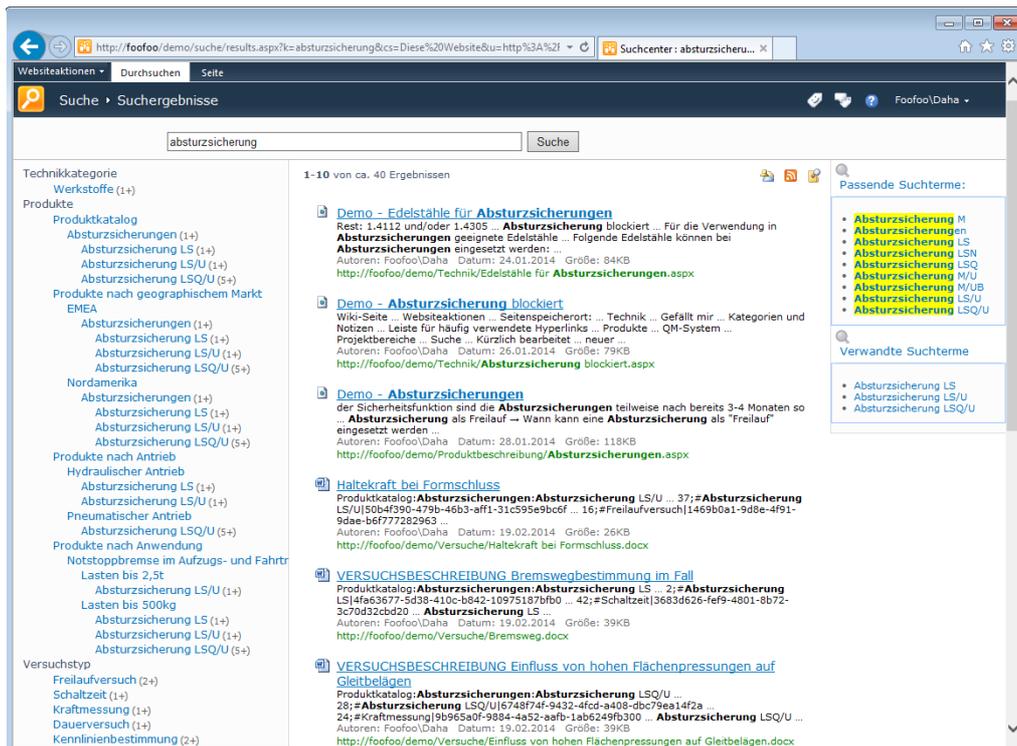


Abb. 1. Ein SharePoint-Suchcenter mit Such-Refinern (links) zur Einschränkung von Suchergebnissen.

Durch Vorliegen einer expliziten Terminologie kann der Suchprozess außerdem besser gesteuert und durch Werkzeuge unterstützt werden. Die Synonymsuche beispielsweise erweitert Suchbegriffe automatisch mit Synonymen und Unterbegriffen, so dass zusätzliche, relevante Suchergebnisse gefunden werden können. Werkzeuge, wie der TermSuggester<sup>5</sup>, unterstützen den Suchprozess dadurch, dass bereits während der Eingabe von Suchbegriffen der TermStore nach passenden Termen durchsucht wird, so dass zielgerichtet nach Dokumenten, die mit diesen Termen annotiert wurden, gesucht werden kann. Wichtig hierbei ist, dass die Suche verbessert wird, ohne dass der Suchende dabei die verwendete Terminologie kennen muss.

SharePoint wird oft als Plattform für die unternehmensweite Suche verwendet, in der unterschiedliche Datenquellen und Unternehmensanwendungen integriert werden (z. B. ERP, CRM, Fileshares oder PLM). Terminologie-Sammlungen können dazu dienen, all diese unterschiedlichen Quellen terminologisch zu integrieren und in einem zentralen SharePoint-Unternehmenssuchcenter mit konsistenten Filtermöglichkeiten zugreifbar zu machen.

<sup>5</sup> [http://www.diqq-pm.de/de/Semantic\\_Search\\_Webparts](http://www.diqq-pm.de/de/Semantic_Search_Webparts)

## Fazit

Eine kuratierte Terminologie enthält branchen- und unternehmensspezifisches Wissen, das die Effektivität und den Erfolg der Suche nach Informationen deutlich verbessern kann. Durch die Verwendung der Termini zur Auszeichnung von Dokumenten und anderen Informationseinheiten werden diese für den erfolgreichen Suchprozess erschlossen. Insbesondere kann die Qualität und Quantität der Metadaten verbessert werden, wenn ein konsistentes Vokabular maschinenlesbar zur Verfügung steht, so dass es z. B. für die automatische Klassifikation von Dokumenten herangezogen werden kann.

Im Gegensatz zu flachen und individuell gepflegten Wortlisten, stellen gepflegte Terminologie-Bestände Synonyme und taxonomisches Wissen bereit. Sie können einem Governance-Prozess unterworfen werden, der ihre Qualität sichergestellt.

Mit der Bereitstellung von gesicherten Terminologie-Listen und Taxonomien kann die Technische Redaktion erhebliche Beiträge zur Verbesserung der Suche in SharePoint leisten. Durch diese Wiederverwendung profitiert das unternehmensweite Wissensmanagement.

## ÜBER RWS Group GmbH

Als Tochter des international agierenden Unternehmens RWS Group (London) ist die RWS Group Deutschland GmbH (Berlin) der kompetente Partner für alle Fragen des effektiven Einsatzes von Sprachtechnologie und der effizienten mehrsprachigen Informationsverwaltung. Wir unterstützen unsere Kunden der Automobil-, Industrie- und Chemiebranche ihre Produkte und Dienstleistungen durch Lokalisierungsmanagement auf internationale Märkte zu vermarkten. Eine zentrale Komponente unserer Dienstleistung ist die Erstellung und Pflege von Glossaren, Taxonomien und Terminologie-Datenbanken und der damit verbundenen Geschäftsprozesse.

Weitere Informationen zum RWS Group Terminologie-Management-Dienstleistungsangebot: <http://rws-group.de/terminologie/>

## ÜBER DIQA PROJEKTMANAGEMENT GMBH

Die DIQA Projektmanagement GmbH aus Karlsruhe ist ein Hersteller von Software-Lösungen für Unternehmensportale in SharePoint und MediaWiki. Mit unseren Lösungen werden Inhalte besser erschlossen, Dokumente sachgerecht bereitgestellt und die Informationssuche beschleunigt. Dazu setzen wir gezielt Taxonomien, Unternehmensontologien und Semantische Technologien ein. Unsere Kunden sind Universitäten, Mittelständler und Großunternehmen aus dem Maschinenbau, Chemie- und Pharma- und Lifesciences-Bereich.

Weitere Informationen zu SharePoint Leistungen und Lösungen von DIQA: <http://www.diqa-pm.de>