

Häufig gestellte Fragen zur PortalU[®]-URL-Pflege

Wie komme ich zur PortalU[®]-URL-Pflege?

Wenn Sie die entsprechende Berechtigung zur Pflege von PortalU[®]-Web- und Katalogseiten haben, haben Sie zwei Möglichkeiten auf die URL-Pflege zuzugreifen.

Möglichkeit 1: Klicken Sie in PortalU[®] (www.portalu.de) auf „Mein PortalU“ und loggen Sie sich hier mit Ihrem Benutzernamen und Passwort ein. Die URL-Pflege (PortalU-Internet-Suchmaschine URL-Pflege) ist unter „iPlugs/iBus“ unterhalb des Reiters „Administration“ zu finden. Hier geben Sie bitte erneut Ihren Benutzernamen und Passwort ein.

Möglichkeit 2: Rufen Sie direkt die URL <http://www.portalu.de/iplug-admin-se-urlmanagement/general/admin-urlmaintenance/index.html> auf und loggen Sie sich hier mit Ihrem PortalU[®]-Benutzernamen und Passwort ein.

Ich bin für das Angebot von **PortalU[®]-Web-URLs** für meine Institution zuständig und habe festgestellt, dass die Anzahl unserer **Webseiten zurückgegangen** ist. Woran könnte das liegen und was kann ich machen, damit unser Angebot wieder besser in PortalU[®] vertreten ist?

Der Rückgang kann verschiedene Ursachen haben:

Eine Änderung der Inhalte und/oder Struktur des Webangebots kann z.B. dazu führen, dass sich die Anzahl der Seiten verringert oder dass bestimmte Teile des Angebots ausgegliedert wurden oder dass die in PortalU eingepflegten URLs veraltet sind.

Ein andere Grund könnte zudem in der Umstellung (September 2010) auf die neue Nutch-Version liegen, da sich mit der neuen Version auch die Technik für die Indexierung geändert hat.

Eine Maßnahme, die in regelmäßigen Abständen (vielleicht alle 6-12 Monate) sinnvoll erscheint, stellt die Überprüfung der Erreichbarkeit der eingestellten Web- und Katalog-URLs im URL-Pflege-Administrationsbereich dar. Außerdem sollte überprüft werden, ob bei den Webseiten der Limitbereich (als der Bereich auf den sich die PortalU-Suche bezieht) richtig gewählt ist.

Wenn in der Spalte "Status" z.B. "moved" bzw. "not found" steht, wurde die Seite verschoben oder existiert nicht mehr.

Eine weitere Maßnahme für die Erhöhung der PortalU-Webseiten besteht in der Recherche weiterer behördlicher Webangebote, die derzeit in PortalU noch nicht erfasst sind.

Eine Start-URL, die ich in die PortalU[®]-URL-Administrationsbereich unter Web-URLs eingepflegt habe, weist den Status „**temp_moved (2010-12-25)**“ aus. Die URL ist in PortalU[®] nicht recherchierbar, lässt sich aber problemlos aufrufen. Woran kann das liegen? Und was muss ich ändern, damit die Seite in PortalU[®] recherchierbar ist?

Der Status "temp_moved (2010-12-25)" besagt, dass die Startseite am 25.12.2010 vom PortalU-Crawler getestet wurde. Der Crawler konnte die Seite nicht fetchen, da die URL auf eine andere URL umgeleitet wird. Um die Webseiten in PortalU wieder sichtbar zu machen, muss die Start-URL entsprechenden angepasst werden.

Eine URL, die ich in die PortalU[®]-URL-Administrationsbereich eingepflegt habe, weist den Status „**not found**“ aus. Die URL ist in PortalU[®] nicht recherchierbar, lässt sich aber problemlos aufrufen. Wenn ich im Admin-Bereich den Test-Button klicke erhalte ich die Information „**Anfrage würde blocken**“. Woran kann das liegen? Und was muss ich ändern, damit die Seite in PortalU[®] recherchierbar ist?

Die Webseiten der Domäne (z.B. www.schleswig-holstein.de) können von PortalU nicht gecrawled werden, da die Domäne dies nicht zulässt. In der entsprechenden **robots.txt** (z.B. www.schleswig-holstein.de/robots.txt) wird PortalU (user-agent: portalu) keine spezifische Crawl-Verzögerungszeit (Crawl-

delay) zugewiesen. Der globale Crawl-delay wurde auf über 30 Sekunden gesetzt. Da PortalU aus Performance-Gründen nach maximal 30 Sekunden aussteigt, wurde die Seite nicht gecrawled. .

Lösungsansatz: Am einfachsten wäre es, die Einträge in der robots.txt wie folgt zu ergänzen:

User-agent: portalU

Crawl-delay: 10

Stand: 24.01.2011