

## Defects Report

## ID (833) - Blokér søgerobotter fra at indeksere undersider

<b>Detected by :</b>	mkkm (Mikkel Mortensen)	<b>Detection date :</b>	27-11-2007
<b>Severity :</b>	3-Medium	<b>Summary :</b>	Blokér søgerobotter fra at indeksere undersider
<b>Actual Fixing time :</b>	39	<b>Category :</b>	Error
<b>Closing Date :</b>	04-01-2008	<b>Closing in version :</b>	
<b>Detection in version :</b>		<b>Estimate (hours) :</b>	
<b>Estimated fixing time :</b>		<b>Fixed in build :</b>	
<b>ID :</b>	833	<b>Modified :</b>	09-01-2008 11:29:19
<b>Planned closing version :</b>		<b>Priority :</b>	
<b>Project :</b>		<b>Reproducible :</b>	Y
<b>Responsible :</b>	mkkm (Mikkel Mortensen)	<b>Status :</b>	Closed
<b>Subject :</b>		<b>System unit :</b>	Retsinformation.dk
<b>Test case number :</b>		<b>ÆA :</b>	

## Comments :

**Mikkel Mortensen <mkkm>, 27-11-2007:** Nu er der implementeret robots.txt-filer til de tre websites. Der benyttes disallow på alt, samt allow på default.aspx og X0500.aspx - siderne. Ikke alle søgemaskiner kan benytte sig af Allow--tag'en, men disse bliver så bare fuldt ud disablede. Problemet med, ikke at gøre dette er, at vi vil være nødt til at vedligeholde robots.txt-filen løbende, såfremt der dukker nye sider op. Med dette fix burde google, msn og yahoo i hvert fald opføre sig efter ønsket. Dette kan dog ikke testes før vi ser fixet på prod, hvorefter det vil fremgå af iis-log'en hvor meget der crawles.

**Birgitte Lund Reichel <blr>, 19-12-2007:** Denne kan ikke testes, før det er implementeret i prod. Efter aftale med SBN tester jeg det på et senere tidspunkt.

**Jeanette Ørnvang <jeanette>, 04-01-2008:** Har søgt på forskrifter og dele indeholdt af forsiderne til LT mv på de tre søgebaser google, msn, yahoo, ok.

## Description :

Efter aftale med cst (Søren), skal der implementeres et filter, således at søgerobotter ikke tillades adgang til samtlige sider på de tre websites. Følgende skal implementeres - robotter skal kun have adgang til følgende:

- forsiderne
- om retsinformation, om lovtidende, om ministerialtidende og de dertilhørende sider

Dermed skal robotter ikke have adgang til sider som ministerieindgangen, abonnementssider, nyeste indhold og selve teksterne.

Ovenstående skal implementeres vha. robots.txt og evt meta-tags.