Finns det några etiska aspekter vid webbskrapning. Kan du hitta något rättsfall?

Det beror på vad man använder informationen till. Skrapning kan användas för att missbruka materialet från originalkällan. Till exempel för att samla in mail adresser och skicka spam. Eller att materialet används för att sätta ihop egna sidor med klick reklam.
Det är viktigt att respektera upphovsrätten, men i många fall är detta ett problem då det inte respekteras.

Finns det några riktlinjer för utvecklare att tänka på om man vill vara "en god skrapare" mot serverägarna?

Att inte belasta servern. Geniom att göra många så kallade requests från servern kan orsaka driftstörningar. Serverägaren kan ha en sämre uppkoppling och kanske därför inte klarar av att hantera för mycket trafik.

Om materialet är upphovsrättsskyddat eller om det klart och tydligt står att serverägaren ber om / förbjuder att skrapa webbsidan så bör man respektera detta. Det går alltid att fråga om tillåtelse. Kanske kan serverägaren göra ett undantag beroende på vad skrapningens syfte är.

Det kan vara bra att identifiera sig själv. Om serverägaren inte tycker om att sidan skrapas så kan skraparen kontaktas om det skulle vara några problem

Begränsningar i din lösning- vad är generellt och vad är inte generellt i din kod?

Så lite hårdkodat som möjligt. Men skrapan är ganska beroende av sidornas innehåll. Det fungerar alltså inte speciellt bra alls om man försöker skrapa någon annan webbsida.
Länkar som hämtas ut har inget beroende av ordning eller värde på href och a tag.

Skrapan är beroende av de tre filmer som finns. Skulle fler filmer läggas till eller någon tas bort på webbsidan så skulle det inte fungera.

Node värdet “ok” som hämtas ut från kalendern kan vara ett nackdel i min skrapa också. Om man byter ut strängen mot något annat så kommer skrapan inte längre att fungera.

Vad kan robots.txt spela för roll?

Det är en fil i webbsidans rotkatalog som identifierar de delar på webbsidan som inte ska vara tillgängliga för sökrobotar. Det kan innebära att serverägaren inte vill att sidan ska besökas av icke människliga användare.