

23. oktober 2024

OB-003c-119b/cbh

Anbefaling om forbehold for tekst- og datamining

Kære medlemmer

I 2023 blev ophavsretsloven ændret pr. 7. juni 2023 og der blev i den forbindelse skabt hjemmel til, at både forskere og alle andre må kopiere ophavsretligt beskyttede værker såsom bøger, hvis det sker til brug for såkaldt tekst- og datamining (TDM). Ved TDM forstås enhver automatiseret, analytisk fremgangsmåde, der har til formål at analysere tekst og data i digital form med henblik på at generere oplysninger, herunder mønstre, tendenser og korrelationer,

Rettighedshavere kan dog tage forbehold mod at andre end forskere må kopiere deres værker til brug for TDM. Da der endnu ikke er national eller EU-retspraksis på området, så er det fortsat tvivlsomt, i hvilken udstrækning TDM-begrebet kan siges at dække over træning af generativ kunstig intelligens.

Kopiering med henblik på TDM forudsætter, at den der kopierer, har en lovlig adgang til indholdet. Det vil f.eks. være hjemmesider, hvor indhold ligger lovligt frit tilgængeligt, og adgang via et abonnement f.eks. til en streamingtjeneste.

Vi har tidligere afholdt et medlemswebinar om TDM-bestemmelserne. Link til webinarret eller til præsentation herfra kan fås ved henvendelse til Rikke rdh@danskeforlag.dk.

I Danske Forlag anbefaler vi, at forlagene hurtigst muligt tager forbehold for TDM. Man kan altid på et senere tidspunkt åbne op igen og evt. gøre en tilladelse betinget af et vederlag.

Hvad skal jeg gøre, hvis jeg vil tage forbehold for TDM?

Vi anbefaler, at forlaget indledningsvist kortlægger, hvor forlagets indhold ligger. Det kan f.eks. være forlagets egen hjemmeside, forhandlere, skoler, eReolen, m.fl.

Indhold, som er tilgængeligt for almenheden

Ift. indhold, som er tilgængeligt for almenheden, skal forbeholdet nedlægges i maskinlæsbar form. Det drejer sig f.eks. om indhold på forlagets egen hjemmeside, indhold på forhandleres hjemmesider m.m.

Den europæiske forlæggerforening FEP har stillet et værktøj til rådighed, som beskriver, hvilke tekniske metoder forlaget kan vælge at benytte sig af for at tage et maskinlæsbart forbehold på web-servers: [Link to the W3C TDM protocol page](#)

Indhold, som ikke er tilgængeligt for almenheden

Her tænkes på indhold, som typisk ligger bag en betalings- eller anden password-beskyttet mur. F.eks. e- og lydbøger hos streamingtjenesterne, indhold hos eReolen og digitale læringsportaler. Her vil der være et aftaleforhold mellem forlaget og den tjeneste, der stiller indholdet til rådighed, samt mellem tjenesten og slutbrugeren.

Hvis forlaget ønsker at tage forbehold for TDM, anbefaler vi, at forlaget i aftalerne med forhandleren skriver, at indhold ikke må benyttes til TDM, og at det af forhandlerens brugerbetingelser ligeledes skal fremgå, at indhold ikke må benyttes til TDM, eller forhandleren nedlægger forbehold i maskinlæsbar form på forhandlerens hjemmeside, jfr. ovenfor.

Forlaget kan f.eks. benytte følgende tekst:

”Det skal fremgå af [forhandlerens] brugerbetingelser, at e-bøger, lydbøger og andet ophavsretligt beskyttet indhold leveret fra [forlaget] til [forhandleren] ikke må benyttes til tekst- og datamining og træning af generative sprogmodeller. I det omfang ophavsretligt beskyttet indhold leveret fra [forlaget] til [forhandleren] ligger åbent tilgængeligt på [forhandlerens] hjemmeside, skal der i maskinlæsbar form tages forbehold for tekst- og datamining. Den europæiske forlæggerforening FEP har stillet et værktøj til rådighed, som beskriver, hvilke tekniske metoder som kan benyttes for at tage et maskinlæsbart forbehold på web-servers: [Link to the W3C TDM protocol page](#)”

Ift. digitale læringsportaler er der tale om web-servers, hvor værktøjet i TDM-protokollen vil kunne benyttes: [Link to the W3C TDM protocol page](#)

Særlige opmærksomhedspunkter

epub

Det er muligt at angive i Pubhub, om der må foretages tekst- og datamining (TDM) på indholdet for både e-bøger og lydbøger. Feltet findes under sektionen "titelinfo".

Alle nye oprettelser står automatisk på "Nej (ingen accept)".

Alle eksisterende titler har fået tildelt "Nej (ingen accept)".

Forlaget kan selv vælge at ændre det til fuld accept. WeDoBooks anbefaler alle at lade den stå på ingen accept for nu.

● Titelformat ^ ✕

Aktiv

Text og data mining (TDM) ⓘ

Notifikationstype ⓘ

ISBN-13 * Record Reference *

Uanset om forlaget har taget forbehold for TDM, må forsknings- og kulturarvsinstitutioner, som driver videnskabelig forskning, foretage TDM, jfr. Indledningen ovenfor.

I ONIX findes TDM i liste 145 (Usage Type), kode 11 (text and data mining).

Bemærk: Tredjepartsmodtagere skal bruge version 11 af API'et for at modtage TDM-metadata. Der er ikke mange, der er på denne version, men WeDoBooks forventer, at alle som minimum har opdateret ifm. at tilgængelighedsloven træder i kraft medio 2025, da de også skal kunne modtage ny tilgængelighedsmetadata.

OpenAI/ChatGPT har introduceret eget forbehold

OpenAI har annonceret, at det er muligt at blokere OpenAIs webcrawler ved at benytte OpenAIs værktøj. Blokeringen virker kun for OpenAIs crawler, den gælder kun fremtidig scraping og forhindrer kun scraping på de hjemmesider, som benytter værktøjet. FEP arbejder for, at tech-virksomhederne ikke opfinder egne forbehold, men at de anerkender W3C's forbehold, som FEP støtter (beskrevet ovenfor). På nuværende tidspunkt er udmeldingen fra FEP dog, at det fuldt ud er muligt at kumulere begge tilgange (ved at bruge opt-out fra OpenAI og W3C-protokollen) for at opnå større sikkerhed.

Link til OpenAI's blokeringsværktøj: [here](#)

Hvis man bruger Cloudflare på hjemmesiden, har de udviklet en funktion, hvor det er muligt nemt at signalere for AI virksomhederne, at man ikke ønsker, at deres crawlere skal kopiere indhold:
<https://blog.cloudflare.com/cloudflare-ai-audit-control-ai-content-crawlers/>

Nota

Danske Forlag har været i dialog med Det Kongelige Bibliotek/Nota, som har meddelt, at Nota pr. 15. juli 2024 på Notas hjemmeside meddeler forbehold i maskinlæsbar form. Ift. brugerbetingelserne oplyser KB/Nota følgende:

"E-bøger og lydbøger, som tilgås/modtages fra Nota, må ikke benyttes til tekst- og data Mining, samt træning af generative sprogmodeller".

Til jeres orientering kan jeg oplyse, at ændringen bliver implementeret flere steder: når et nyt medlem melder sig ind i Nota, får medlemmet et velkomstbrev, hvori den ønskede tekst om forbehold

fremover vil indgå. Derudover skal medlemmet acceptere "Vilkår for brug af materiale", hvori den ønskede tekst også vil fremgå. Disse ændringer er nu implementeret med virkning fra i dag.

Efter et års medlemskab skal hvert enkelt medlem genbekræfte sin accept af vilkår. Også i denne genbekræftelse er der ændret i teksten, så i den sammenhæng sker implementering løbende med fuld implementering i løbet af et år. "

Forsknings- og kulturarvsinstitutioner

Uanset forbehold må forsknings- og kulturarvsinstitutioner, som driver videnskabelig forskning, foretage TDM, hvis de har en lovlig adgang til værket. Institutionerne skal derfor betale for et abonnement, hvis værket kun er tilgængeligt bag en betalingsmur.

Ved en **forskningsorganisation** forstås et universitet, herunder dets biblioteker, et forskningsinstitut eller enhver anden enhed, hvis primære mål er at udføre videnskabelig forskning eller udføre uddannelsesaktiviteter, der også omfatter videnskabelig forskning, som er organiseret uden fortjeneste for øje, eller som investerer hele overskuddet i sin videnskabelige forskning eller varetager af medlemsstaten anerkendte alment nyttige opgaver. Dette på en sådan måde, at adgangen til de resultater, der frembringes af den pågældende enhed af en sådan videnskabelig forskning, ikke kan nydes på et privilegeret grundlag af en virksomhed, som udøver afgørende indflydelse på en sådan institution.

Ved en **kulturarvsinstitution** forstås et offentligt tilgængeligt bibliotek eller museum, et arkiv eller en film- eller lydarvsinstitution.

Christine Bødtcher-Hansen