

Detaljplaner Digitalisering av planbeskrivningar

DiSa – Digital samhällsbyggnad
Datum 2020-05-11

Bakgrund	1
Användningsområden.....	1
Teknisk lösning.....	2
Genomförande av OCR-tolkning.....	2
1) Extern leverantör genomför OCR-tolkningen	2
2) Extern konsult genomför OCR-tolkningen.....	3
3) Region Gotland genomför OCR-tolkningen	3
Resultat	3
Fortsatt arbete.....	5

Bakgrund

I DiSa-projektet identifierade vi ett behov av att, utöver plankartor, också digitalisera de till detaljplanerna hörande planbeskrivningarna. Dessa fanns endast tillgängliga i ett så kallat digitiserat format, som inskannade PDF-filer. I detta format är de dock inte sök- eller maskinläsbara vilket innebär ett hinder för fortsatt digitalisering och verksamhetsutveckling. I DiSa har vi därför tittat på möjligheterna att konvertera planbeskrivningarna till ett mer digitalt format. Det rör sig om cirka 850 filer, varav 150 stycken redan är skapade i moderna programvaror.

Användningsområden

I Plan- och bygglagens kunskapsbank står det att ”en planbeskrivning är ett verktyg som ska ge intressenter, sakägare och tillsynsmyndigheter en bild av vad detaljplanen innebär, hur den påverkar intressen utifrån olika aspekter och vilka avvägningar som gjorts mellan olika intressen.” Planbeskrivningarnas innehåll regleras dels i Plan- och bygglagen (PBL) och dels i Boverkets föreskrifter och allmänna råd. Visst innehåll, som planens syfte och genomförande, är obligatoriskt i alla planbeskrivningar medan annat innehåll bara krävs i vissa planer.

[Webbplats för Plan- och bygglagens kunskapsbank \(öppnas i nytt fönster\)](#)

En viktig effekt av att digitalisera planbeskrivningarna är att allmänheten kan göra sökningar efter ord eller fraser i dokumenten. Planbeskrivningarna är generellt inte längre än 20 sidor långa (med flertalet bilder) så att leta manuellt i ett dokument tar inte alltför lång

tid. Undantag finns dock, till exempel planbeskrivningen för Bunge Biskops 1:23 som är på 111 sidor, eftersom den innehåller ett gestaltungsprogram. Att leta i flera filer sparar också mycket tid med hjälp av textsökning.

Ett praktiskt exempel på användning av planbeskrivningar som dykt upp i projektet är att förrättningslantmätare har användning för detta. De använder bland annat beskrivningar för att kontrollera särskilda bestämmelser som inte tydligt förklaras i plankartan. Planens syfte, som är juridiskt bindande, är också något som de använder sig av i förrättningsärenden.

För att kunna fortsätta digitaliseringen av planbeskrivningar krävs också ett annat format än de bildfiler som äldre, inskannade, beskrivningar lagras som. Föreskrifterna för hur den vidare digitaliseringen ska ske är vid DiSa-projektets slut 2021 inte helt klara. Lantmäteriet och Boverket har dock intentionen att texter från planbeskrivningarna ska knytas till geografiska områden i plankartan. Texten om planens syfte knyts till hela planområdet och texter om enskilda bestämmelser till respektive bestämmelseyta. Det kommer inte att bli lagkrav på att genomföra denna digitalisering för redan existerande planer men för att få samma funktionalitet i dessa som de nya, helt digitala, planerna krävs någon form av insats.

Teknisk lösning

Den vanligast förekommande tekniken för att omvandla inskannade textfiler till textfiler är så kallad OCR-tolkning, där OCR står för Optical Character Recognition eller optisk teckenigenkänning. Det är en metod i vilken ett program optiskt konverterar bilder till textfiler. Metoden är automatisk och förhållandevis pålitlig. Den har dock sina begränsningar, framförallt vad gäller handskriven text, text som har redigerats i efterhand (till exempel som fått ett streck över sig) och filer som har skannats in med sämre kvalitet. På grund av detta kan den ursprungliga tolkningen behöva kontrolleras efter OCR-tolkningen.

Genomförande av OCR-tolkning

I DiSa har vi undersökt tre huvudspår för att utföra tolkningen.

1. Extern leverantör genomför OCR-tolkningen
2. Konsultfirma genomför OCR-tolkningen
3. Region Gotland genomför OCR-tolkningen

1) Extern leverantör genomför OCR-tolkningen

Vi hittade och kontaktade fyra stycken möjliga leverantörer. Två stycken meddelade att de inte kan erbjuda den här typen av tjänst. Då återstod två leverantörer.

Leverantör 1

Cirka 20 kronor per fil

Två testfiler ingår

Totalt: 16 000 kronor med metadata, 3 500 kronor utan metadata

I detta ingår OCR-tolkning, specificerade metadata som XML-fil, omdöpta filnamn (om det behövs) och en sökbar Excel-databas. Programvaran som används är IRIS Powerscan 10.

Efter vidare kontakt med leverantören konstaterades att Region Gotland inte har behov av tillhörande metadata. Detta då planbeskrivningarna redan har namn som knyts till planerna och det i programvaran för detaljplaner (GEOSECMA Fysisk Planering) går att bland annat få ut ålder på respektive plan. Med detta blev erbjudna pris 3 500 kronor.

Leverantör 2

Startavgift: 300 kr

Löpande: 6 kr/fil

Totalt: 4 500 kr

I detta ingår kvalitetssäkring genom stickprovskontroller på vissa sökord. Ytterligare jämförelse mot original kan göras mot timplenning. Leverantören erbjöd också en test-tolkning av en fil. Programvaran som används är Adobe Acrobat.

2) Extern konsult genomför OCR-tolkningen

En konsultfirma tar hela ansvaret för att göra tolkningen. Detta skulle förslagsvis göras genom programvaran Adobe Acrobat Pro. Tidsåtgången för tolkning och kvalitetssäkring är svår att uppskatta. Med 10 minuter/fil skulle tidsåtgången då bli cirka 120 timmar och totalkostnaden 120 timmar gånger konsultens timpris. Timpriset kan hållas nere genom att praktikanter gör större delen av jobbet.

3) Region Gotland genomför OCR-tolkningen

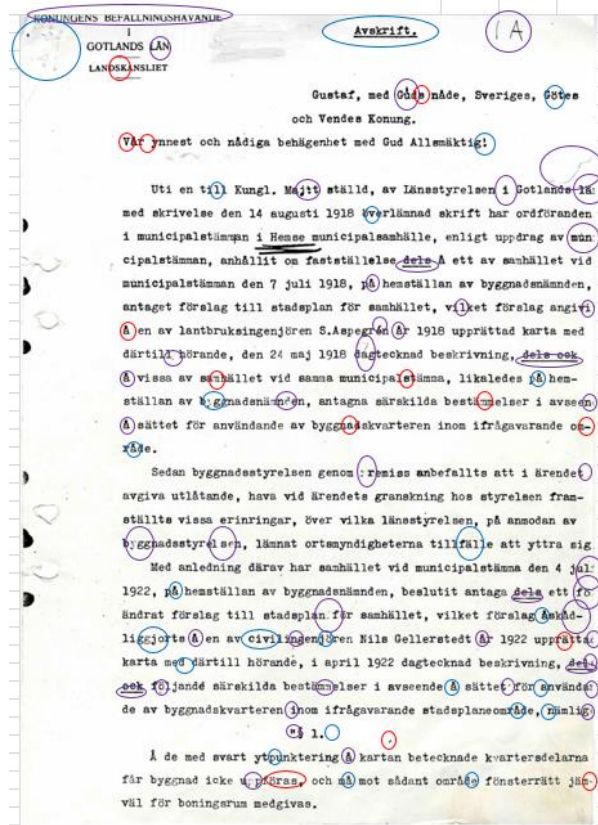
Mycket likt alternativ 2, men kostnaden blir i form av arbetstid snarare än fakturerade kronor.

Resultat

I DiSa-projektet valde vi en extern leverantör då en helautomatisk OCR-tolkning av en leverantör ansågs mest kostnadseffektivt och tillräckligt för att höja användbarheten på nuvarande planbeskrivningar. Knappt 700 filer, den äldsta från 1923 och den nyaste från 2005, skickades till båda leverantörerna. Detta för att vi var osäkra på om resultatet skiljde mycket och vi ville därför kunna jämföra resultatet sinsemellan. De tolkade filerna levererades efter någon arbetsdag.

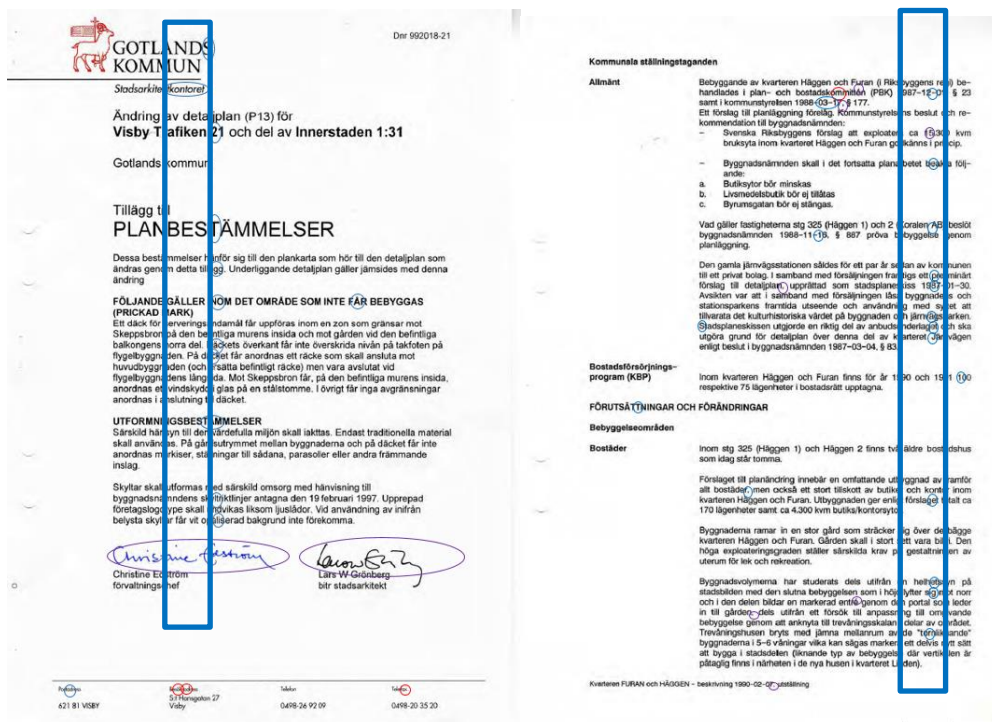
Utan möjligheter att göra någon form av automatisk uppföljning av hur väl programmen lyckades med tolkningen är det svårt att dra några långgående slutsatser. Stickprovskontroller visar dock att:

- Det finns ett antal generella tolkningsproblem med särskilda tecken som mellan
 - m och n
 - i och 1 och l
 - å och ä
- Formatering, som understreck, fet och kursiv stil, och accenter inte kommer med
- Bindestreck som ligger i slutet på en rad slår ihop ordet med nästa ord vilket ger lustiga effekter på ord, som Restaurang- och hotellverksamhet eller Gatu- och parkmark



Figur: En slumpvis utvald sida från en OCR-tolkad planbeskrivning. Röda markeringar är fel i den ena leverantörens tolkning, blåa markeringar fel i den andra leverantörens tolkning och lila markeringar fel i båda tolkningarna. Notera särskilt att all text inte fått plats till höger på sidan och att många fel därför uppstått där.

Vi såg inga större skillnader i resultaten mellan de två leverantörerna och programvarorna. En leverantörs filer har visat tendenser till att ha fel som inte har någon direkt koppling till ”svårare” tecken eller skador på pappret. Istället verkar dessa fel finnas på en viss del av sidan trots att det inte finns några synliga problem med det området och i den andra leverantörens tolkning fanns inte samma problem. Se två exempel på detta nedan. Något som däremot är tydligt är att kvaliteten på inskanningen spelar roll. I de fall där all text inte får plats på sidan blir det till exempel ökar antalet fel markant. Övergången från skrivmaskins- till datorskrivna dokument verkar ha skett runt 1990 och efter detta blir det färre fel i tolkningarna.



Figur: Två sidor där OCR-tolkningen resulterat i fel i ett område på sidan där inga synliga skador finns.

Beslutet blev att de OCR-tolkade filerna från leverantör 1 fick, utan korrigeringar, ersätta de planbeskrivningar som allmänheten hade tillgång till via webbkartorna på gotland.se/kartor. När detta gjordes upptäcktes snabbt att plankartorna, som varit förstasida i de tidigare filerna, saknades. Planbeskrivningarna som skickades på tolkning innehöll inte dessa vilket flera personer, både medarbetare på Region Gotland och allmänheten hörde av sig om. Då det skulle ha tagit lång tid att sammanfoga de OCR-tolkade planbeskrivningarna med respektive plankarta manuellt, och ingen automatisk metod för detta upptäcktes, skickades nya filer där plankartan ingick från start för ny OCR-tolkning. På så sätt finns nu OCR-tolkade planbeskrivningar med plankartor tillgängliga både intern och externt.

När dessa filer publicerades noterade dock medarbetare att proportionerna i vissa kartor blivit felaktiga i tolkningen. Då det kan vara relevant att mäta i originalritningarna, till exempel om någon form av tvist uppstår, är det viktigt att skalor är korrekta. Leverantören gjorde därför en ny leverans av filer där detta fel enligt dem inte längre fanns. Stickprovskontroller visade också att kartorna nu hade samma proportioner som filerna som skickades över.

Fortsatt arbete

Om planbeskrivningarna ska digitaliseras ytterligare, till exempel enligt Boverkets kommande specifikation, duger inte den textkvalitet som OCR-tolkningen resulterade i. Någon form av korrigeringsarbete måste i sådant fall göras. En möjlig efterarbetsmetod vore att söka efter vissa skiljetecken, som: ~ ; : ! • /

Dessa tecken förekommer normalt inte särskilt ofta i en planbeskrivning men ingår ofta i tolkningarnas resultat. På så sätt kan arbetet med att hitta fel och rätta till dem minskas något.

För att kunna göra en digitalisering av planbeskrivningar i stor skala, även efter att textkvaliteten höjs, krävs också bättre verktyg. Ett systemstöd, likt det som finns för plankartorna, effektiviserar förmodligen processen mycket. Det är också intentionen att GEOSECMA Fysisk planering ska stödja även kommande specifikationer vilket gör att systemet kan användas även för detta.

I ett korrigerings- och digitaliseringsarbete kan det vara bra att prioritera vissa planbeskrivningar. En möjlig prioritering är då att börja med beskrivningar från och med 1987 års plan- och bygglag. Detta då dessa beskrivningar innehåller mer information. Dessa är också framförallt skrivna på dator vilket innebär att det finns färre fel i tolkningen. Ett annat alternativ är att logga och följa upp antalet sidvisningar på de olika dokumenten för att sedan påbörja arbetet med de dokument som använts mest. Ett tredje alternativ är att låta användarna redigera dokumenten och att de kan bidra med rättningar till tolkningen. I båda de senare alternativen finns dock tekniska utmaningar som i sådant fall behöver lösas.