

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Informaatika õppekava

Risto Rahulan  
Automaatne sisukokkuvõtete tegemine ajaseoste  
märgenduste põhjal

Bakalaureusetöö (9 EAP)

Juhendaja: Siim Orasmaa, PhD

Tartu 2021

## **Automaatne sisukokkuvõtete tegemine ajaseoste märgenduste põhjal**

**Lühikokkuvõte:** Sisukokkuvõtted on eriti olulised tänapäeval, kus informatsiooni on väga palju. See aitab inimestel kõige tähtsama osa tekstist kiiresti kätte saada. Hea lahendus oleks luua tekstidest automaatselt sisukokkuvõtted. Selles töös kasutatakse Eesti TimeML korpuse märgendatud artikleid, et sündmuste ajaseoste märgendite põhjal luua ajagraafid ja kasutada neid automaatselt artiklite kokkuvõtete tegemisel. Töö käigus hinnatakse kokkuvõtete kvaliteeti kasutades ROUGE-L meetodit.

**Võtmesõnad:** Sisukokkuvõtja, sisukokkuvõtja hindamine, ajaseosed, ajagraafid, TimeML, EstTimeMLCorpus

**CERCS:** P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

## **Automatic Summarization Based on Temporal Semantic Annotations**

**Abstract:** Summaries are especially important today, where there is a lot of information. This helps to get the most important parts of the text quickly. A good solution would be to automatically create summaries for the texts. In this work, annotated articles from the Estonian TimeML corpus are used to create temporal relation graphs based on the temporal relations of events. These graphs are used to automatically create summaries from the articles. The quality of the automatically created summaries is evaluated using the ROUGE-L method.

**Keywords:** Summarization, summarization evaluation, temporal relations, temporal relation graph, TimeML, EstTimeMLCorpus

**CERCS:** P170 Computer science, numerical analysis, systems, control

<b>1. Sissejuhatus</b>	<b>4</b>
<b>2. Automaatsed sisukokkuvõtjad</b>	<b>5</b>
2.1 Eestikeelsed sisukokkuvõtjad	5
2.2 Ingliskeelsed sisukokkuvõtjad	5
2.3 Automaatsete sisukokkuvõtjate hindamine (ROUGE-L)	6
2.4 TimeML märgendusraamistik	7
2.4 Ajasemantiliste märgendustega tekstikorpus (EstTimeMLCorpus)	7
<b>3. Eeltöö</b>	<b>9</b>
3.1 Arendus- ja testkorpus	9
3.3 Märgendamine	9
<b>4. Sisukokkuvõtete tegemine ajaseoste märgenduste põhjal</b>	<b>10</b>
4.1 Ajagraafid	10
4.2 Ajagraafide edasiarendus	12
<b>5. Tulemused</b>	<b>15</b>
5.1 Tulemuste analüüs	15
5.2 Tulevikusuund	17
<b>6. Kokkuvõtte</b>	<b>18</b>
<b>Viidatud kirjandus</b>	<b>19</b>
<b>Lisad</b>	<b>20</b>
I. Artikkel “aja_sloleht_2007_12_17.tasak.a003”	20
II. Artikkel “aja_ml_2002_47.tasak.a006”	21
III. Kirjutatud kood ja referentskokkuvõtted	22
IV. Litsents	23

## 1. Sissejuhatus

Sisukokkuvõtted on eriti olulised tänapäeval, kus informatsiooni on väga palju. See aitab inimestel kõige tähtsama osa tekstist kiiresti kätte saada. Samuti aitavad sisukokkuvõtted otsustada, kas antud tekst sisaldab just seda, mida otsiti ja kas seda on mõtet läbi lugeda. Sisukokkuvõtete loomine automaatselt on siinjuures hea lahendus. Kokkuvõtteid on võimalik luua ülevaate (ingl abstract) ja väljavõtte (ingl extractive) meetodeid kasutades [1]. Ülevaate meetodi puhul luuakse kokkuvõttesse uued laused, mis erinevad algtekstist kuid annavad edasi sama sisu [1]. Väljavõtte meetodi puhul valitakse laused algtekstist ja neid ei muudeta [1]. Automaatne sisukokkuvõtja, mis kasutab väljavõtte meetodit, otsustab, millised teksti osad on olulised ja eemaldab ülejäänud muu. Siit tuleb ka probleem, millised laused tekstis annavad edasi põhiideed ja kuidas neid lauseid üldse välja valida. Selle jaoks on mitu lähenemisvõimalust, näiteks statistiline, teemapõhine ja graafipõhine [1]. Käesolevas töös kasutatakse kokkuvõtte loomisel graafipõhist väljavõtte meetodit, kasutades tekstis olevaid sündmusi, ajaväljendid ja seoseid nende vahel. Antud töö eesmärk on kasutada ajasemantilisi märgendusi, et luua automaatselt tekstist sisukokkuvõte ja hinnata loodud kokkuvõtete kvaliteeti.

Käesolev töö koosneb neljast osast. Esimeses osas antakse ülevaade mõnest olemasolevast kokkuvõtte loomise lahendusest, loodud kokkuvõtete kvaliteedi hindamisest. Samuti kirjeldatakse töös kasutusel olevaid andmeid. Teises osas antakse ülevaade arendus- ja testkorpusest ja teksti eeltööst. Kolmandas osas kirjeldatakse meetodeid, mida kasutati automaatselt kokkuvõtete loomisel. Neljandas osas hinnatakse loodud kokkuvõtete kvaliteeti ja võrreldakse neid EstSum tulemustega.

## 2. Automaatsed sisukokkuvõtjad

Selles peatükis antakse ülevaade olemasolevatest automaatsetest sisukokkuvõtjatest. Kirjeldatakse, milliseid andmeid kasutatakse selle töö jooksul ja kuidas hinnata automaatselt loodud sisukokkuvõtte kvaliteeti.

### 2.1 Eestikeelsed sisukokkuvõtjad

Eesti keele jaoks on olemas erinevaid lahendusi, kuidas automaatselt luua tekstist kokkuvõtte. Kaili Müürisep [2] on kirjeldanud eestikeelset sisukokkuvõtjat EstSum, mis loob põhiliselt uudiste artiklite põhjal kokkuvõtteid. EstSum leiab kokkuvõtte tegemisel lausetele skoori sõnavara, asukohta ja formaadi abil. EstSum hindab lauseid, kus esinevad paksus kirjas sõnad kõrgemalt ja lausetele, kus on jutumärgid antakse madalam skoor. Leiti, et artikli esimene lause esines tekstis iga kord. Keili Sellik [3] on hinnanud EstSumi poolt loodud kokkuvõtteid ja ka sealt tuleb välja, et artikli esimene lause on kõige olulisem. Janar Saks [4] on lisanud EstSumile lingvistilise mooduli EstNLTK, mis võtab lausete analüüsimisel arvesse sõnade algvormi. Eesmärk oli selle abil saada lausetest rohkem informatsiooni ja tulemuse hindamisel saadi loodud kokkuvõttest testkorpuse peal 2.17% parem tulemus ja arenduskorpuse peal oli tulemus 4.16% parem. Saks toob ka välja, et tulemus on saadud väikse korpuse peal ja suurema korpuse peal testides saaks teha üldisemaid järeldusi.

### 2.2 Ingliseelsed sisukokkuvõtjad

Tabak ja Evrim [5] käsitlevad artiklis sündmusepõhist sisukokkuvõtte loomist uudiste artiklitest. Allikas on välja toodud, et sündmused, mis on sõltuvad või sisaldavad teisi sündmusi on omavahel seotud semantiliselt. Kõigepealt tehakse teksti eeltöötlus, siis leitakse sündmuste tunnused ja valitakse sealt välja sobivad. Siis valitakse laused, kusjuures alati võetakse esimene lause, sest see sisaldab olulist informatsiooni. Ülejäänud lausete puhul kontrollitakse tunnuste jada olemasolu ja selle abil hinnatakse, kui oluline on antud lause tekstis. Nii valitaksegi laused kokkuvõtte jaoks.

Eesti keele peal ei ole proovitud kasutada ajasemantika märgendusi sisukokkuvõtete loomisel. Inglise keele on seda tehtud. Maofu Liu jt [6] on kirjutanud automaatse sisukokkuvõtte loomisest kasutades tegusõnade semantilist suhet. Täpsemalt vaadeldi "toimub enne" (ingl *happens-before*) ajalisi suhteid. Mingile tegevusele eelneb mingi teine tegevus. Näiteks, enne millegi kritiseerimist seatakse miski kahtluse alla. Kasutati tegusõnade võrgustikku, kus on kirjeldatud tegusõnade omavahelised ajalised suhted. Neid suhteid on võimalik väljendada graafina. Teksti põhjal luuakse graaf ja sellest graafist leitakse kõige pikem ahel. Eeldati, et see ahel graafis on seotud tekstis kõige olulisema informatsiooniga. Lõpuks vaadeldakse tekstis olevaid lauseid, mis sisaldavad ahelas olevaid tegevusi ja nende hulgast valitakse välja need, mis sobiksid kokkuvõttesse. Otsuse tegemisel hinnatakse tegevuse olulisust ja kui sama tegevus esineb mitmes lauses, siis valitakse ainult üks neist, sest sama tegevusega on seotud sarnane informatsiooni.

Tekstide automaatse sisukokkuvõtte loomisel annab häid tulemusi teksti formaadi arvesse võtmine kui ka teksti esimese lause sisaldamine kokkuvõttes [2]. Samuti annab tulemust parandada rakendades sõnadele lemmatiseerijat [4]. Ka ajaseoste kasutamine kokkuvõtte loomisel on lootusrikas, millele siinne töö keskendubki.

### 2.3 Automaatsete sisukokkuvõtjate hindamine (ROUGE-L)

Kuidas hinnata automaatselt loodud sisukokkuvõtete kvaliteeti? Üks võimalus oleks lugeda loodud kokkuvõtet ja otsutada, kas see annab edasi olulist informatsiooni ja kas loodud tekst on sidus ja hästi loetav. Selline lähenemine võtaks kaua aega ja tulemused ei oleks mõõdetavad. Selleks, et hinnata automaatselt loodud kokkuvõtte kvaliteeti oleks seda vaja võrrelda sellise kokkuvõttega, mis on hea [3]. Referentskokkuvõtted on käsitsi koostatud ja sisaldavad algtekstist väljavalitud lauseid [3]. Varasemalt on eesti keelel toimiva sisukokkuvõtja EstSum hindamisel võrreldud automaatselt loodud kokkuvõtte kattuvust referentskokkuvõtetega [4]. Samuti on EstSumi hindamisel kasutatud Rouge paketti [3].

ROUGE (ingl *Recall-Oriented Understudy for Gisting Evaluation*) on viis hinnata automaatselt loodud kokkuvõtete kvaliteeti [7]. See võimaldab võrrelda automaatselt loodud sisukokkuvõtet referentskokkuvõttega nelja erineva meetodiga: Rouge-N, Rouge-S, Rouge-L, Rouge-W [7].

Rouge-N mõõdab sõnade järjendeid pikkusega  $n$  ehk  $n$ -grammide kattuvust automaatselt ja käsitsi loodud kokkuvõtetes [7].

Rouge-S leiab sellised bigrammid, kus sõnad peavad olema samas järjestuses, kuid ei pea olema kõrvuti [7]. Näiteks lauses “Eile toimus neli avariid” on 6 sellist bigrammi: “Eile toimus”, “Eile neli”, “Eile avariid”, “toimus neli”, “toimus avariid”, “neli avariid”. Rouge-S mõõdab selliste bigrammide kattuvust hinnatava kokkuvõtte ja referentskokkuvõtte vahel.

Rouge-L puhul leitakse pikim ühisjada  $LCS(X,Y)$  (*Longest Common Subsequence*) kus  $X$  laused referentskokkuvõttes ja  $Y$  on laused kandidaat kokkuvõttes [7]. Mida pikem on kahe teksti vahel leitud ühisjada, seda sarnasemad on ka kokkuvõtted [7]. Näiteks on lausete jada  $X = [A,C,E]$ ,  $Y_1 = [A,BC]$  ja  $Y_2 = [A,D,F]$ , siis pikimad ühisjadad oleks  $LCS(X,Y_1) = [A,C]$  ja  $LCS(X,Y_2)=[A]$ . Kuna  $LCS(X,Y_1)$  ühisjada on pikem kui  $LCS(X,Y_2)$  siis ka  $Y_1$  antakse parem tulemus.

Rouge-W on täiustatud versioon Rouge-L meetodist. Siin salvestatakse leitud ühisjadade järjestikuste vastete pikkused [7]. See parandab Rouge-L tulemust eelistades sellist pikimat ühisjada, milles rohkem järjestikulisi vasteid [7]. Kui on jada  $X=[\underline{A} \underline{B} C D]$  ja jada  $Y_1 = [\underline{A} F L \underline{B}]$  ning  $Y_2 = [\underline{A} \underline{B} L F]$ . Leitakse pikim ühisjada  $LCS(X,Y_1)$  ja  $LCS(X,Y_2)$ . Mõlemal juhul pikima ühisjada pikkus on 2, kuid kuna  $Y_2$  puhul on leitud pikim ühisjada elemendid järjestikku, siis oleks see parem valik.

Käesolevas töös kasutati hindamiseks Rouge-L meetodit, sest sisukokkuvõtete loomisel on kasutatud väljavõtte meetodit ehk kokkuvõtetes on muutmata kujul laused algtekstist.

## 2.4 TimeML märgendusraamistik

Alljärgnev materjal on refereeritud Pustejovsky jt [8] konverentsartiklist.

TimeML on märgenduskeel, kus sündmused ja ajaväljendid ning nende omavahelised seosed on tekstis esile toodud. TimeML sisaldab nelja tüüpi märgendeid. EVENT tüüpi märgendit kasutatakse sündmust tähistavate sõnade puhul. EVENT tüüpi märgendid on jagatud seitsmesse erinevasse klassi:

Juhtum (*occurrence*) - Näiteks sõnad: ehitama, müüma, liituma.

Seisund (*state*) - Näiteks sõnad: on, kuulub, veendunud

Teavitust (*reporting*) - Näiteks sõnad: ütleva, teatama

Etteavatsetud tegevus (*I-Action*) - Näiteks sõnad: proovima, lubama, pakkuma

Etteavatsetud seisund (*I-State*) - Näiteks sõnad: uskuma, tahtma, kavatsema

Aspektiline (*Aspectual*) - Näiteks sõnad: algama, lõppema, jätkama

Taju (*Perception*) - Näiteks sõnad: nägema, kuulma, vaatama

TIMEX3 on ajaväljendite märgendamiseks. Ajaväljenditeks võivad olla kindlad kuupäevad ja kellajad. Ajaväljendid võivad olla ka ebatäpsed, näiteks “homme”, “eelmine aasta”, “kaks päeva tagasi”. TIMEX3 märgendit kasutatakse ka kestvuse märgendamiseks (“Kaks aastat”, “Viis kuud”).

SIGNAL tüüpi märgendit kasutatakse selliste sõnade märgendamiseks, mis viitavad ajaliste seoste. Näiteks sõnad: enne, pärast, samal ajal. SIGNAL märgendi juurde käib ka ajaliste seoste eitused (mitte enne, mitte samal ajal) kui ka ajaväljendi toimumise arv (kaks korda, kolm korda).

LINK märgendeid kasutatakse sündmuste ja ajaväljendite vaheliste seoste märkimiseks. On kolme tüüpi LINK märgendeid: TLINK, SLINK ja ALINK. TLINK märgendi abil tähistatakse seosed sündmuste või sündmuste ja ajaväljendite vahel. SLINK märgendiga on märgendatud seosed kahe sündmuse või sündmuse ja SIGNAL märgendi vahel. ALINK väljendab seoseid aspektiliste sündmuste ja selle argument sündmuse vahel.

## 2.4 Ajasemantiliste märgendustega tekstikorpuse (EstTimeMLCorpus)

Selle töö juures on kasutusel Eesti TimeML märgendatud korpus, mis kasutab TimeML märgendusraamistikku. Korpus on kättesaadav GitHubis<sup>1</sup>. See sisaldab 18 artiklit ajalehest “Maaleht”, 36 artiklit ajalehest “Postimees” ja 26 artiklit ajalehest “Õhtuleht”. Keskmiselt on artiklis 17 lauset ja keskmiselt igas lauses 16 sõna. Korpus on käsitsi korrastatud ja sellele on lisatud ajamärgendused. Korpuses on märgendatud sündmused, ajaväljendid ja seosed nende vahel. Seal on üheksa erinevat TLINK märgendi seoste tüüpi.

Selged seosed on:

- BEFORE - ajaliselt eelnev
- AFTER - ajaliselt järgnev
- SIMULTANEOUS - ajaline kattuvus
- IDENTITY - ajaline kattuvus ja tegemist sama sündmusega

---

<sup>1</sup> <https://github.com/soras/EstTimeMLCorpus>

- IS\_INCLUDED - ajaline kaasamine (A on Bs)
- INCLUDES - ajaline kaasamine (B sisaldab Ad)

Ebaselged seosed on:

- BEFORE-OR-OVERLAP - ajaliselt enne või kattuv
- OVERLAP-OR-AFTER - ajalist kattuv või järgnev
- VAGUE - Ebaselge ajaline seos

Näide märgenditest ühes lauses:

“70 mln kr maksev sadam [tahetakse ] valmis [saada ] [2004. aasta maiks ]”

*[SAADA] BEFORE-OR-OVERLAP [2004. aasta maiks]*

Sadam tahetakse valmis saada hiljemalt maiks.

*[SAADA] AFTER [DOKUMENDI LOOMISE AEG]*

Sadam saadakse valmis pärast artikli kirjutamist.

*[TAHETAKSE] BEFORE OR OVERLAP [SAADA]*

Kõigepealt tahetakse saada valmis sadam. Tahtmine pädib tahetud asja juhtumisega, ehk siin saadakse sadam valmis.

Ajaseoste märgendus on jagatud neljaks kihiks. Esimeses kihis on ajaseosed sündmuse ja ajaväljendite vahel. Teises kihis on ajaseosed sündmuste ja dokumenti loomise kuupäeva vahel. Kolmandas kihis on ajaseosed peassündmusete vahel kõrvuti olevates lausetes. Neljandas kihis on ajaseosed samas lauses olevate sündmuste vahel. Täpsem kirjeldus korpusest on saadaval Eesti Keeleressursside Keskuse veebilehel<sup>2</sup>.

---

<sup>2</sup> <https://keeleressurssid.ee/et/276-ajasemantiliste-margendustega-tekstikorpus>

### 3. Eeltöö

Selles peatükis kirjeldatakse arendus- ja testkorpus. Selle loomist ja eeltööd, selleks et rakendada EstSum kokkuvõtjat võrdluseks.

#### 3.1 Arendus- ja testkorpus

Korpus jagati arendus- ja testhulgaks, kus arendushulgas oli 64 artiklit ja testhulgas 16 artiklit. Korpus viidi tekstilisele kujule, mida oleks kerge lugeda. Korpuses puuduvad kokkuvõtted iga artikli kohta ja need tuli käsitsi ise luua. Iga korpuses oleva artikli juures valiti välja laused, mis sobivad kokkuvõttesse. Iga artikli kohta loodi üks referentskokkuvõtte. Võrreldes esialgsete tekstidega, on kokkuvõtte pikkus umbes 70% lühem. Kokkuvõtete jaoks lausete valimisel põhiline probleem seisnes selles, et vahepeal oli keeruline valida, millised laused sobiksid kõige paremini. Selgelt oli näha, et tihti sobivad mitmed laused ja milline lause on tähtsam kokkuvõtte jaoks on suures osas subjektiivne. Näiteks laused artiklist aja\_sloleht\_2007\_12\_17.tasak.a003 (Lisa 1):

1. Kirurg : abielu oli võimatu
2. Sunday Telegraph kirjutab, et Hasnat Khan lõpetas suhted Dianaga 1997. aasta suvel, mõni kuu enne printsessi surma , sest tema veendumuse kohaselt oluaks nende abielu võimatu."
3. Kui ma võtaksin ta naiseks, siis lahutaksime hiljemalt aasta pärast," oli Hasnat oma perekonnale öelnud."
4. Mu poeg mõtles paluda tema kätt, kuid loobus siis sellest, sest teadis, et abielu kahe eri usku inimese vahel ei saa olla püsiv," ütles isa.

Need laused edastavad sama ideed, kui kuna kokkuvõtte pikkus on piiratud, siis peaks nende lausete hulgast valida ühe. Autori poolt loodud referentskokkuvõtted on saadaval GitHubis<sup>3</sup>.

#### 3.3 Märgendamine

Selleks, et testida ja võrrelda ajasemantiliste märgendite abil automaatselt loodud teksti kokkuvõtteid EstSumiga tuli Eesti TimeML korpuses olevad artiklid viia sobivale formaadile, mida on kirjeldatud EstSumi Github repositooriumis<sup>4</sup>. Täpsemalt oli vaja korpus joondada originaaltekstiga ja lisada vastavalt algtekstidele tekstivormingu märgendused. Pealkirjad on `<h1></h1>` märgendite vahel, lõigud `<p></p>` vahel ja laused `<s></s>` vahel. Kui tekst on paksus kirjas või kaldkirjas siis lisatakse see osa tekstist vastavalt `<hi rend="bold"></hi>` või `<hi rend="italic"></hi>` vahele. Pealkirjad ja kõik alampealkirjad on märgendite `<head></head>` vahel. Oluline on ka see, et iga lause oleks eraldi real ja lõikude algus ja lõpp märgid oleksid samuti eraldi ridadel. Neid märgendeid saaks ka kasutada automaatse sisukokkuvõtja juures lisaks ajasemantilistele märgenditele, mitte ainult tulemuste võrdlemisel EstSumiga.

<sup>3</sup> Koodi repositoorium: <https://github.com/Risto-R/AutomaatneSisukokkuvotja>

<sup>4</sup> EstSum Github: [https://github.com/janarsaks/EstSum\\_development](https://github.com/janarsaks/EstSum_development)

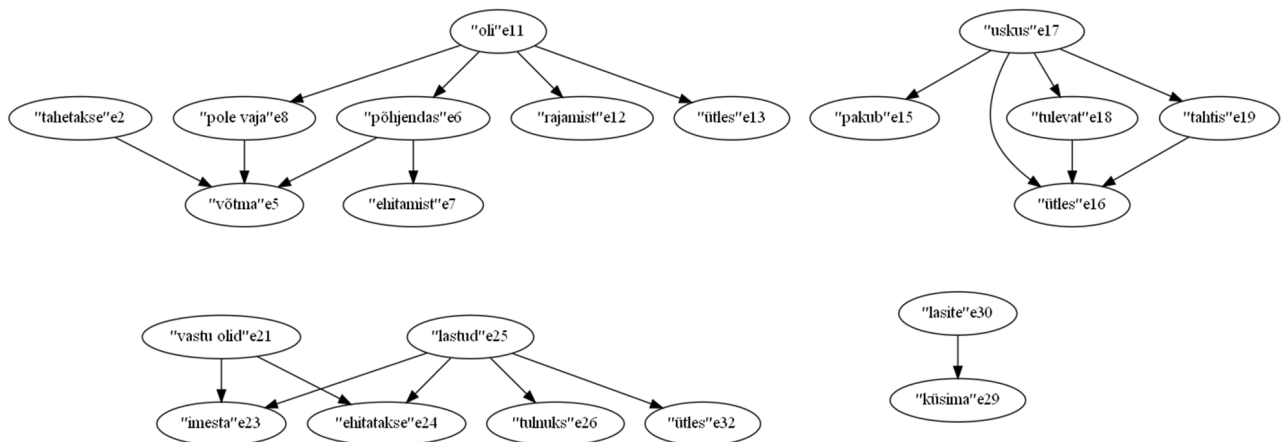
## 4. Sisukokkuvõtete tegemine ajaseoste märgenduste põhjal

Selles peatükis kirjeldatakse ajagraafe ja kuidas nende põhjal loodi automaatselt sisukokkuvõtteid. Antakse ka hinnang meetodide tulemusele.

### 4.1 Ajagraafid

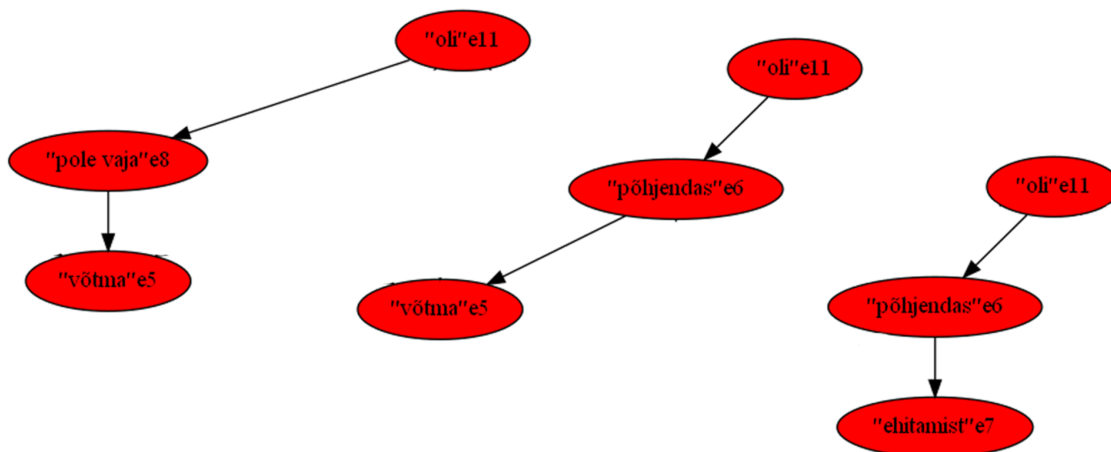
Maofu Liu jt [6] on kirjeldanud, kuidas luua etteantud tekstist ajagraaf ja kuidas neid kasutada automaatselt sisukokkuvõtte loomisel. Idee on luua suunatud graaf sündmuste vahel, kus lähtetipp on sündmus, mis toimub enne lõpptippul olevat sündmust.

Joonisel 1 on loodud graaf ühest korpuses olevast artiklist (Lisa 2).



Joonis 1. Ajagraaf artikli “aja\_ml\_2002\_47.tasak.a006.sol” põhjal.

Kuna kõik sündmused tekstis ei ole üldjuhul seotud, siis saadud graaf ei ole täisgraaf. Graafist leitakse maksimaalne nõrgalt sidus komponent ehk selline sidus komponent, kus on kõige rohkem tippu, sest just see võiks viidata teksti peamisele teemale. Joonisel 1 olevas graafis on sellises komponendis kaheksa tippu. Võib juhtuda, et leitud komponendis on tsüklid, sellisel juhul tuleb need eemaldada, et ei tekiks lõpumatut tsüklit hiljem, kui graafist otsitakse pikimat ahelat. Tsükli esinemisel eemaldatakse tsüklist see kaar, mille lõpptippul on suurim sisendaste. Selleks, et ajagraafi kasutada sisukokkuvõtte tegemiseks, tuleks selle abil leida kõige olulisem informatsioon tekstis. Seda saab teha, leides pikima ahela maksimaalsest nõrgalt sidusast komponendist. Mõte on selles, et pikim ahel graafi maksimaalses nõrgalt sidusast komponendist näitab teksti peamist sündmuste käiku. Kui on mitu pikimat ahelat, siis valitakse esimene. Joonisel 2 on kujutatud graafist leitud pikimad ahelad.



Joonis 2. Maksimaalne nõrgalt sidusa komponendi pikimad ahelad artiklist  
 “aja\_ml\_2002\_47.tasak.a006.sol”.

Peale pikima ahela leidmist, saab seda kasutada kokkuvõtte loomisel. Vaatluse all on need laused, kus esinevad ahelas leiduvad sündmused. Kuna saadud lausete hulk võib olla suurem kui kokkuvõttesse lubatud, siis tuleb hinnata lausete tähtsust tekstis ja lisada kokkuvõttesse ettenähtud mahus lauseid tähtsuse järgi. Üks võimalus panna lausele numbriline väärtus on vaadata lausetes leiduvate sündmuste astet algses ajagraafis, ehk sündmustega seotud kaarte arvu, ja liita need kokku. Laused, mille sündmuste aste on suurem on ka tähtsamad, kuna need sündmused on seotud rohkem teiste sündmustega ülejäänud tekstis. Kui kokkuvõtteks on laused välja valitud, tuleks need panna samasse järjekorda, nagu olid need originaalses tekstis, selleks et laused oleksid loogilises järjekorras ja sidusamad.

PageRank on rekursiivne graafi põhine algoritm, mis hindab graafis olevate tippude olulisust [9]. See loodi, et hinnata veebilehtede olulisust ja järjestada need otsingumootoris [9]. Siin saab seda kasutada, selleks et hinnata lausete olulisust teksti põhjal loodud graafist. Maofu Liu jt [6] on kirjeldanud ka PageRank algoritmi kasutamist kokkuvõtte loomisel. Luuakse samamoodi, nagu eelnevalt, suunatud graaf sündmuste vahel, kus lähtetipp on sündmus, mis toimub enne lõpptipul olevat sündmust. Siis kasutatakse PageRank algoritmi, et hinnata tippude olulisust.

Eelnevalt kirjeldatud algoritmi kasutati Eesti TimeML korpuse peal. Kasutati korpuses kõiki AFTER ja BEFORE seoseid, et luua graaf. Leiti pikim ahel graafi maksimaalses nõrgalt sidusas komponendist ja hinnati lausete olulisust tippude astmete järgi. Samuti hinnati tippude olulisust PageRank algoritmi abil. Kuna käesoleva töö puhul loodud kokkuvõtetel on pikkuse piirang (umbes 30% algteksti pikkusest), siis järjestati laused vastavalt lausetes olevate sündmuste olulisuse hinnangule ja vajadusel eemaldati madalama hinnanguga laused. Tabelis 1 on arendushulga ROUGE-L tulemused.

	Saagis ( <i>recall</i> )	Täpsus ( <i>precision</i> )	F-mõõt ( <i>F-measure</i> )
Tippude aste	0.24873	0.44595	0.30314
PageRank	0.46927	0.45784	0.45890

Tabel 1. Maofu Liu jt [6] eeskujul loodud kokkuvõtete ROUGE-L tulemused.

Kui loodi automaatselt kokkuvõtted maksimaalse nõrgalt sidusa komponendi pikima ahela abil, siis saadi ROUGE-L saagiseks 0.24873. PageRank abil kokkuvõtte loomisel saadi 88.6% parem tulemus. Pikima ahela meetodi puhul on lausete valik piiratud, vahel leitakse ainult üks lause kokkuvõttesse. PageRanki kasutades on kõik teksti laused järjestatud lause olulisuse järgi ja ka valik, milliseid lauseid kokkuvõttesse pannakse, on suurem. See on üks põhjus, miks PageRank tulemus on kordades parem.

## 4.2 Ajagraafide edasiarendus

Eelnevalt kasutati ajagraafide loomisel ainult ajalisel eelnevaid ja ajalisel järgnevaid seoseid. Eesti TimeML korpuses on olemas veel erinevaid seoseid sündmuste vahel (vt 2.4). Samuti pikima ahela leidmisel maksimaalsest nõrgalt sidusast komponendist, valiti ainult üks pikim ahel, kuid neid võib olla mitu. Üks võimalus oleks kasutada kõiki sama pikkusega pikimaid ahelaid. Sellisel juhul oli kokkuvõtte loomisel ROUGE-L saagis 0.28379, täpsus 0.44365 ja F-mõõt 0.33071. Võrreldes ainult ühe pikima ahela valikuga oli tulemus 14% parem.

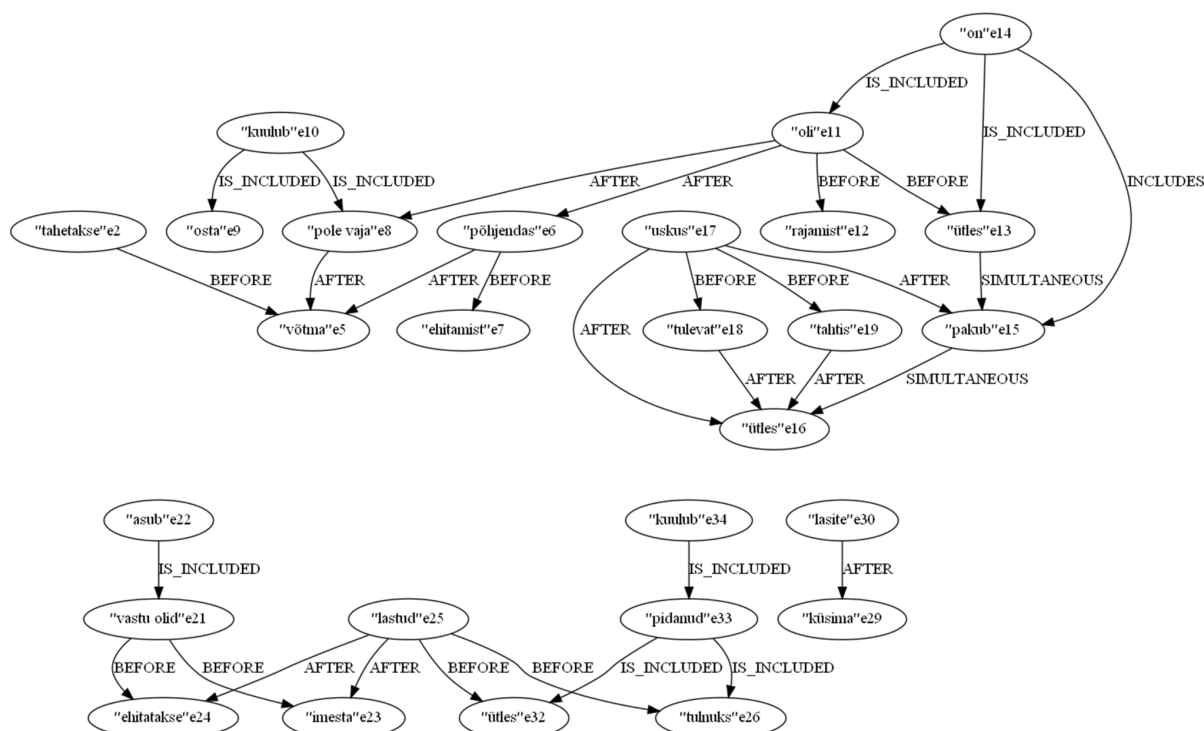
Ajagraafidesse lisati madalama kaaluga ebaselged seosed BEFORE-OVERLAP ja AFTER-OVERLAP. Parima tulemuse andis, kui nende seoste kaal oli 0.4. Pikima ahela leidmise meetodiga oli ROUGE-L saagis 0.23460, täpsus 0.38149 ja F-mõõt 0.27738. Tulemus oli 5.6% kehvem, kui kasutati ainult selgeid seoseid. Kui pikima ahela puhul võtta arvesse ka teisi sama pikkusega ahelaid, siis oli saagis 0.25516, täpsus 0.39491 ja F-mõõt 0.29715. Ka see tulemus on kehvem, 10% võrra. Ka PageRank andis halvema tulemuse graafis, mis sisaldab ebaselgeid seoseid. PageRank ROUGE-L saagis 0.46356, täpsus 0.44594 ja F-mõõt 0.45065. Siit võib järeldada, et ebaselgeid seoseid ei ole mõtet kasutada, sest automaatsete sisukokkuvõtete kvaliteet on madalam. Samuti on tulemus parem, kui võtta arvesse kõiki pikimaid ahelaid graafi maksimaalses nõrgalt sidusas komponendis, mitte ainult ühte neist.

Järgnevalt prooviti kasutada Eesti TimeML korpuses olevaid teisi sündmuste vahelisi seoseid. Parim tulemus saavutati, kui kasutati sündmuste vahelisi seoseid, kus sündmus A on lähtetipp ja sündmuse B on lõpptipp:

- A BEFORE B
- B AFTER A
- A SIMULTANEOUS B

- B IS\_INCLUDED A
- A INCLUDES B
- A IDENTITY. B

Joonisel 3 on kujutatud artikli “aja\_ml\_2002\_47.tasak.a006.sol”(Lisa 2) põhjal loodud graafi. Joonisel 3 oleval graafil näitavad seoste suunda nooled ja kaare juures on esialgne märgend. Graafist leiti maksimaalne nõrgalt sidus komponent ja sealt pikimad ahelad. Tipu astmete järgi lausete hindamisel saavutati loodud kokkuvõtetel ROUGE-L saagis 0.32764, täpsus 0.42348, F-mõõt 0.35893. Sellisel graafil PageRank algoritmi abil sisukokkuvõtete loomisel oli saagis 0.45165, täpsus 0.41234, F-mõõt 0.42807. Pikima ahela meetodi puhul tulemus paranes võrreldes eelmistega. PageRank tulemus oli kehvem, kuid siiski parem, kui pikima ahela abil sisukokkuvõtte loomine.



Joonis 3. artikli “aja\_ml\_2002\_47.tasak.a006.sol” põhjal loodud graafi.

Korpuses on ka olemas ajaväljendid ja nende seosed sündmustega. Prooviti lausete olulisuse hindamisel lisaks tipu astmetele võtta arvesse ka ajaväljendite olemasolu. Nendele lausetele, mis sisaldasid ajaväljendeid lisati olulisuse skoorile 0.3. Tulemuseks saadud sisukokkuvõtete ROUGE-L tulemus paranes. Saagis oli 0.340001, täpsus oli 0.44555, F-mõõt oli 0.37473.

	Kasutatud seosed	Saagis ( <i>recall</i> )	Täpsus ( <i>precision</i> )	F-mõõt ( <i>F-measure</i> )
Pikim ahel (Üks pikim ahel)	BEFORE, AFTER	0.24873	0.44595	0.30314
PageRank	BEFORE, AFTER	0.46927	0.45784	0.45890
Pikim ahel (kõik pikimad ahelad)	BEFORE, AFTER	0.28379	0.44365	0.33071
Pikim ahel (Üks pikim ahel)	BEFORE, AFTER, BEFORE-OVERLAP, AFTER-OVERLAP	0.23460	0.38149	0.27738
PageRank	BEFORE, AFTER, BEFORE-OVERLAP, AFTER-OVERLAP	0.46356	0.44594	0.45065
Pikim ahel (kõik pikimad ahelad)	BEFORE, AFTER, BEFORE-OVERLAP, AFTER-OVERLAP	0.25516	0.39491	0.29715
Pikim ahel(kõik pikimad ahelad)	BEFORE, AFTER, SIMULTANEOUS, IS_INCLUDED, INCLUDES, IDENTITY	0.32764	0.42348	0.35893
PageRank	BEFORE, AFTER, SIMULTANEOUS, IS_INCLUDED, INCLUDES, IDENTITY	0.45165	0.41234	0.42807
Pikim ahel(kõik pikimad ahelad, ajaväljendite olemasolu)	BEFORE, AFTER, SIMULTANEOUS, IS_INCLUDED, INCLUDES, IDENTITY	0.340001	0.44555	0.37473

Tabel 2. Ajagraafide meetodite ROUGE-L tulemused arenduskorpusel.

Tabelist 2 on näha, et parimad tulemused saavutati, kui kasutati PageRank algoritmi graafis, mis sisaldab ainult selgeid BEFORE ja AFTER seoseid sündmuste vahel. Samuti saavutati hea ROUGE-L hinnang sisukokkuvõtetele, mille loomisel on kasutatud maksimaalse nõrgalt sidusast komponendist pikimaid ahelaid graafist, mis sisaldab BEFORE, AFTER, SIMULTANEOUS, IS\_INCLUDED, INCLUDES, IDENTITY seoseid sündmuste vahel, samuti võeti arvesse ajaväljendite olemasolu lausetes.

## 5. Tulemused

Selles peatükis hinnatakse parimaid ajagraafide meetodeid testkorpusel. Kasutatakse eestikeelset sisukokkuvõtjat EstSum, et võrrelda tulemusi.

### 5.1 Tulemuste analüüs

Kahte parimat meetodit hinnati testkorpusel. PageRank tulemus testkorpusel on parem, kui arenduskorpusel. Pikima ahela meetodi puhul on tulemus halvem. Ka testkorpusel on PageRank algoritmi kasutades graafist sisukokkuvõtte jaoks lausete valimine andnud parima tulemuse. Käesolevas töös kasutati EstSum sisukokkuvõtjat võrdluse eesmärgil. EstSumi jaoks joondati korpuses olevad tekstid algtekstidega, kus on olemas tekstivormingu märgendid. EstSumi kasutati vaikeväärtustega ehk positsiooni skoori osakaal oli 0.4, formaadi skoori osakaal 0.4 ja võtmesõnade skoori osakaal 0.2.

	Kasutatud seosed	Saagis ( <i>recall</i> )	Täpsus ( <i>precision</i> )	F-mõõt ( <i>F-measure</i> )
Pikim ahel(kõik pikimad ahelad, ajaväljendite olemasolu)	BEFORE, AFTER, SIMULTANEOUS, IS_INCLUDED, INCLUDES, IDENTITY	0.30556	0.45902	0.35996
PageRank	BEFORE, AFTER	0.48125	0.48140	0.46705
EstSum		0.47611	0.60431	0.52534

Tabel 3. Ajagraafide meetodite ROUGE-L tulemused testkorpusel ja EstSumi tulemused.

EstSumi F-mõõt on 12.48% parem PageRank meetodi omast. Kuid kuna sisukokkuvõtte pikkus oli mõlemal juhul piiratud, siis on kasulik vaadata saagist. Ajagraafist PageRank algoritmi poolt loodud sisukokkuvõtte saagis on 1.08% parem. Oluline on märkida, et saadud tulemused on subjektiivsed, kuna referentskokkuvõtted on loodud ainult autori poolt. Kuna kokkuvõttesse võib valida erinevaid laused, mis on sisult sarnased, siis tuleks vaadata, millised on loodud sisukokkuvõtted päriselt. Artikli “aja\_ml\_2002\_47.tasak.a006.sol” (Lisa 2) põhjal loodud kõigi kolme meetodi sisukokkuvõtted:

Pikima ahela meetod:

5 - Tammel oli kalasadam veel 1960. aastate lõpul , enne uue kalasadama rajamist Tagalahte Veerele .

6 - Saare maavanem Jüri Saar ütles , et looduslikult on Tamme ilusaim sadamakoht nii maalt kui merelt vaadatuna .

7 - Maavanem pakub selle Põhja-Saaremaale ehitatava sadama nimeks selguse mõttes Saaremaa sadam .

8 - Mustjala vallavanem Enno Kolter ütles , et uskus juba kaheksa aastat tagasi sadama tulevat just Tammele .

PageRank meetod:

3 - Hooaja jooksul peaks sadam vastu võtma 30-35 kruisilaeva .

4 - Tallinna Sadama juhtkond põhjendas sadama ehitamist Tammele väikseimate süvenduskuludega , samuti pole vaja sadama maa-ala

8 - Mustjala vallavanem Enno Kolter ütles , et uskus juba kaheksa aastat tagasi sadama tulevat just Tammele .

13 - Varem või hiljem tulnuks see sadam teha , sest muidu hakatakse küsima , miks te ühe projekti põhja lasite , aga asemele midagi ei pakkunud , " ütles Lokk .

EstSum:

1 - Tallinna Sadam ehitab Saaremaa süvasadama Küdema lahe läänerannikule endisesse Tamme sadamakohta Mustjala valda .

7 - Maavanem pakub selle Põhja-Saaremaale ehitatava sadama nimeks selguse mõttes Saaremaa sadam .

9 - Uudepanga lahte tahtis süvasadamat ehitada juba 2000. aasta suveks AS Undva Sadam .

11 - Toonane AS Undva Sadama nõukogu esimees Jaak Lokk ei imesta , et süvasadam nüüd lõpuks ikkagi ehitatakse .

Referentskokkuvõtte:

1 - Tallinna Sadam ehitab Saaremaa süvasadama Küdema lahe läänerannikule endisesse Tamme sadamakohta Mustjala valda .

2 - 70 mln kr maksev sadam tahetakse valmis saada 2004. aasta maiks .

3 - Hooaja jooksul peaks sadam vastu võtma 30-35 kruisilaeva .

4 - Tallinna Sadama juhtkond põhjendas sadama ehitamist Tammele väikseimate süvenduskuludega , samuti pole vaja sadama maa-ala kelleltki välja osta , sest see kuulub riigile .

Referentskokkuvõttesse on pandud neli esimest lauset. Pikima ahela meetodil loodud sisukokkuvõttes ei ole ühtegi kattuvat lauset. Samas, artikli peamine mõte, et ehitatakse sadam, on seal siiski olemas. PageRank kokkuvõttes kattuvad referentskokkuvõttega pooled laused ja EstSumi puhul kattub ka ainult üks lause. Pikima ahela meetodi puhul on selle artikli sisukokkuvõttes kõik laused järjestikku aga üldiselt on seda meetodit kasutades ka teiste artiklite puhul sisukokkuvõttesse pandud järjestikused laused. See teeb sisukokkuvõtte paremini loetavamaks Samas tähendab see seda, et kui artikli alguses olevates lausetes, mis annavad edasi kõige rohkem olulist infot, ei ole sidusad laused, siis neid ka kokkuvõttesse ei valita. Nii EstSumi ja PageRank põhjal loodud sisukokkuvõtted on sobivad, mõlemad annavad edasi artikli peamise idee, kõrvaline info on neil erinev. Kumb neist on kvaliteetsem sisukokkuvõtte, on subjektiivne.

## 5.2 Tulevikusuund

ROUGE-L tulemuste põhjal võib väita, et ajaseoste kasutamine automaatselt sisukokkuvõtete loomisel on teostatav ja annab kvaliteetseid tulemusi. Selles töös oli kasutusel ainult autori poolt loodud referentskokkuvõtted. Objektiivsema tulemuse jaoks oleks vaja mitut referentskokkuvõtete komplekti, mis on loodud erinevate inimeste poolt. Samuti võiks uurida, kui edukas on ajaseoste abil sisukokkuvõtete loomine teistsuguste tekstide puhul, eriti selliste tekstide juures, mis ei ole nii selgelt struktureeritud nagu on uudiste artiklid. Tulevikus ajaseoste rakendamisel sisukokkuvõtte loomisel võiks ka rakendada teisi meetodeid ajagraafide kõrval, et valida välja lausest oluline informatsioon ja ajagraafide abil hoida kokkuvõtte sidus. Näiteks kombineerida EstSumi lähenemine ja ajagraafid.

## 6. Kokkuvõtte

Käesolevas töös kasutati automaatselt sisukokkuvõtte loomisel ajaseoste märgendusi. Eesti TimeML korpusel olevates uudiste artiklites on märgendatud erinevate sündmuste vahelised seosed. Esmalt loodi korpusel olevate tekstide põhjal sisukokkuvõtted, et neid hiljem võrrelda automaatselt loodud sisukokkuvõtetega. Töö käigus loodi sündmuste seoste põhjal ajagraafid. Ajagraafist leiti maksimaalne nõrgalt sidus komponent, ning sellest pikim ahel. Just see peaks andma edasi teksti peamist ideed. Parim tulemus saadi, kui pikimas ahelas olevate lausete olulisust hinnati ajagraafi tippude astme ja ajaväljendite olemasolu alusel. Sisukokkuvõtte kvaliteeti hindamisel saadi ROUGE-L saagiseks 0.340001, täpsuseks 0.44555 ja F-mõõt oli 0.37473. Sellest parema tulemuse andis ajagraafide peal PageRank algoritmi rakendamine. PageRank hindas ajagraafis olevate sündmuste olulisust ja selle põhjal valiti välja ka laused. ROUGE-L saagiseks 0.48125, täpsuseks 0.48140 ja F-mõõt oli 0.46705. Ajaseoste märgenduste põhjal automaatselt sisukokkuvõtetete tegemine on paljulubav ja andis häid tulemusi. Objektiivsema hinnangu andmiseks tuleks luua mitu komplekti referentskokkuvõtteid ja rakendada ajagraafe suurema hulga tekstide peal.

## Viidatud kirjandus

- [1] Gambhir, M., Gupta, V. Recent automatic text summarization techniques: a survey. *Artif Intell Rev* 47, 2017, 1–66.  
<https://doi.org/10.1007/s10462-016-9475-9><https://link.springer.com/article/10.1007/s10462-016-9475-9> (01.08.2021)
- [2] Müürisep, Kaili. Eestikeelsete tekstide sisukokkuvõtjast EstSum. *Kogumikus Keel ja arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised* 6.(Toim. M. Koit, R. Pajusalu, H. Õim) Tartu 2006. lk 115-125. <http://uuslepo.it.da.ut.ee/~kaili/papers/myyrisepsisukokku.pdf> (10.01.2021).
- [3] Sellik, Keili. 2008. “Automaatse sisukokkuvõtja töö hindamine”. TÜ arvutiteaduse instituudi bakalaureusetöö. [http://lepo.it.da.ut.ee/~kaili/juhendamised/Baka\\_Sellik.pdf](http://lepo.it.da.ut.ee/~kaili/juhendamised/Baka_Sellik.pdf) (10.12.2020).
- [4] Saks, Janar. 2018. “Eestikeelsete tekstide sisukokkuvõtja EstSum edasiarendamine”. Tü arvutiteaduse instituudi bakalaureusetöö.  
[https://comserv.cs.ut.ee/ati\\_thesis/datasheet.php?id=61941&year=2018](https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=61941&year=2018) (10.12.2020).
- [5] Tabak, Feride Savaroğlu; Evrim, Vesile, “Event-based summarization of news articles”, *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol. 28 Issue 2, 2020, p850-864. 15p, 850-857 (10.12.2020)
- [6] Maofu Liu, Wenjie Li, Huijun Hu. “Extractive Summarization Based on Event Term Temporal Relation Graph and Critical Chain”. *Information Retrieval Technology*. Springer Berlin Heidelberg, 2009, 87-99.  
[https://link.springer.com/chapter/10.1007/978-3-642-04769-5\\_8](https://link.springer.com/chapter/10.1007/978-3-642-04769-5_8) (10.12.2020).
- [7] Lin, C.Y.. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, Association for Computational Linguistics, 2004, (pp. 74–81).  
<https://aclanthology.org/W04-1013.pdf> (01.07.2021).
- [8] Pustejovsky, James & Castaño, José & Ingria, Robert & Saurí, Roser & Gaizauskas, Rob & Setzer, Andrea & Katz, Graham & Radev, Dragomir. *TimeML: Robust Specification of Event and Temporal Expressions in Text.. New Directions in Question Answering*. 2003, 28-34.  
[https://www.researchgate.net/publication/221441154\\_TimeML\\_Robust\\_Specification\\_of\\_Event\\_and\\_Temporal\\_Expressions\\_in\\_Text](https://www.researchgate.net/publication/221441154_TimeML_Robust_Specification_of_Event_and_Temporal_Expressions_in_Text) (01.08.2021).
- [9] Page L, Brin S, Motwani R, Winograd T. *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab; 1999.  
<http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf> (04.08.2021)

## Lisad

### I. Artikkel "aja\_sloleht\_2007\_12\_17.tasak.a003"

"Printsess Diana süda oli valust lõhkemas , kui Hasnat Khan teatas , et ei abiellu temaga . Ta otsis Dodi Fayedilt vaid lohutust , mitte armastust , " kinnitas Diana ja Dodi hukkimise asjaolusid uurivas Londoni kohtus printsessi lähemaid sõbratare Rosa Monckton . " Diana armastas kirurg Hasnat Khani kogu südamest ja lootis , et neil on ühine tulevik , " kinnitas Rosa Monckton kohtule Sunday Telegraphi andmeil . Kõigest kaks nädalat enne oma traagilist surma , Kreeka saartel puhates , oli Diana rääkinud sõpradega ainult Hasnatist . " Ma mõistsin , et ta igatseb seda meest . Kordagi polnud jutuks abiellu Dodi Fayediga , " ütles Monckton . Printsess oli valmis vastu võtma islami . See samm jäi astumata vaid sellepärast , et Hasnat otsustas nende vahekorrale lõpu teha . " Kui ajalehtedes ilmusid esimesed fotod Dianast ja Dodist , oli printsess ehmunud ja endast väljas ning ütles mulle : " Nüüd ei võta Hasnat mind enam tagasi . " " Kirurg : abiellu oli võimatu

Sunday Telegraph kirjutab , et Hasnat Khan lõpetas suhted Dianaga 1997. aasta suvel , mõni kuu enne printsessi surma , sest tema veendumuse kohaselt oluaks nende abiellu võimatu . " Kui ma võtaksin ta naiseks , siis lahutaksime hiljemalt aasta pärast , " oli Hasnat oma perekonnale öelnud . " Me oleme kultuuritaustalt erinevad , nagu võiksid olla inimesed kahelt eri planeedilt . Tema on Veenuselt , mina Marsilt . " Kirurgi ja printsess Diana vahekord algas 1995. aastal . Enamasti viibisid armunud Hasnat Khani korteris Londoni Chelseas . Tõsi , Diana varjas seda suhet avalikkuse eest suure hoolega ja tema armastusest Pakistani päritolu kirurgi vastu oli ajakirjanduses tema eluajal vähe juttu . Praegu 48aastane Hasnat lahkus Suurbritanniast vaid paar kuud tagasi . Tema isa rääkis ajakirjanikele , et poeg oli Dianat iseloomustanud kui väga iseseisvat ja lahke olekuga naist . " Mu poeg mõtles paluda tema kätt , kuid loobus siis sellest , sest teadis , et abiellu kahe eri usku inimese vahel ei saa olla püsiv , " ütles isa . Hasnat Khan abiellus enne Suurbritanniast lahkumist Afganistani kuningasuguvõsa liikme , 29aastase Hadia Sher Aliga . Rosa Moncktoni jutu kohtus pareeris Diana kasuema krahvinna Raine Spencer , kes ütles , et Dodi Fayed ja Diana olid armunud , ja on usutav , et neil oli plaan abielluda . Kuid krahvinna polnud printsessile kuigi lähedane . Diana karske suhe Dodiga

Seevastu on printsessi lähedaste sõprade hulka loetud ajakirjanik Taki Theodoracopoulos varem kinnitanud , et Diana ja Dodi ei maganud kunagi ühes voodis . " Praegu kõneldakse palju , et kui Diana Alma silla tunnelis surma sai , siis ootas ta Dodi last . Ma küsisin seda temalt kolm päeva enne õnnetust - ta naeris mulle näkku ja ütles , et kui Dodi on tema armuke , siis olen seda ka mina . Nad ei maganud koos , " ütles Taki . Londoni kohus on Diana ja Dodi Fayedide surma asjaolude kohta tunnistusi kuulates muu hulgas välja nõudnud printsessi kirjad Dodile ja kuninganna Elizabeth II abikaasale Philipile .

## II. Artikkel "aja\_ml\_2002\_47.tasak.a006"

20.11.2002

Tallinna Sadam ehitab Saaremaa süvasadama Küdema lahe läänerannikule endisesse Tamme sadamakohta Mustjala valda . 70 mln kr maksev sadam tahetakse valmis saada 2004. aasta maiks . Hooaja jooksul peaks sadam vastu võtma 30-35 kruisilaeva .Tallinna Sadama juhtkond põhjendas sadama ehitamist Tammele väikseimate süvenduskuludega , samuti pole vaja sadama maa-ala kellelki välja osta , sest see kuulub riigile .Tammel oli kalasadam veel 1960. aastate lõpul , enne uue kalasadama rajamist Tagalahte Veerele .Saare maavanem Jüri Saar ütles , et looduslikult on Tamme ilusaim sadamakoht nii maalt kui merelt vaadatuna Maavanem pakub selle Põhja-Saaremaale ehitatava sadama nimeks selguse mõttes Saaremaa sadam .Mustjala vallavanem Enno Kolter ütles , et uskus juba kaheksa aastat tagasi sadama tulevat just Tammele .Uudepanga lahte tahtis süvasadamat ehitada juba 2000. aasta suveks AS Undva Sadam .Selle vastu olid aga eelkõige linnuteadlased , sest Uudepanga lahes asub ohustatud linnu kirjuhaha suurim talvitusala Läänemeres .Toonane AS Undva Sadama nõukogu esimees Jaak Lokk ei imesta , et süvasadam nüüd lõpuks ikkagi ehitatakse . " Džinn oli ju pudelist välja lastud . Varem või hiljem tulnuks see sadam teha , sest muidu hakatakse küsima , miks te ühe projekti põhja lasite , aga asemele midagi ei pakkunud , " ütles Lokk . Ornitoloogid on pidanud parimaks sadamakohaks Veeret , mis praegu kuulub AS Hiiu Kalurile.

### **III. Kirjutatud kood ja referentskokkuvõtted**

Autori poolt loodud referentskokkuvõtted ja kood sisukokkuvõtete loomiseks on saadaval GitHub repositooriumis <https://github.com/Risto-R/AutomaatneSisukokkuvotja>

## **IV. Litsents**

### **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Risto Rahulan,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Automaatne sisukokkuvõtete tegemine ajaseoste märgenduste põhjal, mille juhendaja on Siim Orasmaa, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Risto Rahulan*

**04.08.2021**