

Commentary/Kommentaar

Filmipildi märksõnastamisest Eesti filmi andmebaasis/ rahvusfilmograafias

Meta-Description of Films in Estonian
Film Database / National Filmography



DE GRUYTER
OPEN

HAGI ŠEIN, Tallinn University / Tallinna Ülikool, Estonian Film Database /
Eesti Filmi Andmebaas; email: hagi@tlu.ee

DOI: 10.1515/bsmr-2015-0007

SISSEJUHATUS

Alates 1912. aastast, mil valmis esimene Eesti film, on Eestis loodud üle 12 000 filmiteose. 2008. aastal alustas MTÜ Eesti Filmi Andmebaas nende filmiteoste universaalse kirjeldussüsteemi loomist info koondamiseks Eesti filmi andmebaasi (EFA), sisuliselt niisiis elektroonilise rahvusfilmograafia tegemist. 2012. aasta lõpus sai andmebaasi esimene järk avalikkusele kättesaadavaks (www.efis.ee).

Artikkel puudutab mõningaid filmiandmete eripära, Eesti filmi andmebaasi metaandmestiku, ennekõike filmiteoste sisu ja tähenduse märksõnastamisega (*subject indexing*) seotud küsimusi ja praktilisi lahendusi, mida andmesüsteemi kavandamise protsessis tuli luua ja arendada. Kuna valdav osa teooriast ja uuringutest pildi märksõnastamise alal käsitleb seisva pildi indekseerimist, vaatleme siin ka olemasolevate kontseptsioonide ja põhimõtete ülekandmise ja rakendamise võimalusi liikuvaid pilte kirjeldavas andmebaasis, tuginedes praktilisele kogemusele.

Filmiandmebaaside üks olulisemaid funktsioone on luua efektiivne ligipääs filmide sisule filmipildi märksõnastamise ja sellel põhineva infootsingu (*information retrieval*) kaudu. Filmide märksõnastamine seisneb sisuliselt pildi sisu ja tähenduse verbaliseerimises. Pildi märksõnastajad nimetavad filmipildis kujutatud objekte, sündmusi, protsesse, tegelasi, suhteid, meeolusid, kohti ja aega filmipilti (ja heli) kirjeldades, analüüsides ja tõlgendades-interpretēerides. Seda tehakse andmebaasi kasutajate huvides, et tagada võimalikult relevantne, asjakohane vastus infootsijat huvitavale küsimusele.

Kuidas analüüsida ja väljendada pildi sisu, milliseid sisukategooriaid ja klassifikatsioone filmide märksõnastamisel kasutada, kuidas seejuures arvestada erinevate filmiliikide ja žanrite spetsiifikat, kas pildi tähenduse indekseerimine peaks olema võimalikult objektiivne ja välistama interpretatsioonid, kuidas minimeerida märksõnastajate erinevatest oskustest ja võimekusest tekkivaid erisusi filmiteose märksõnastiku tähendusruumis, mida

reaalselt tähendab filmide kirjeldamise protsessis nõue arvestada potentsiaalsete kasutajate huvide ja vajadustega? Need on audiovisuaalteoste katalogiseerijate-indekseerijate ringis jätkuvalt aktuaalsed küsimused, mille üle püüan Eesti filmi andmebaasi loomise kogemusele toetudes järgnevas arutleda.

1. EESTI FILMI ANDMEBAAS

1.1. Eesmärgid

Eesti mängufilmid, dokumentaal- ja aimefilmid, anima-, tele-, õppe-, reklaam- ja kroonikafilmid on rikkalik kogu Eesti elust, ajaloost, kultuurist ja inimestest. Eesti filmide loomisel on osalenud ligi 3000 filmitegijat, valdav osa Eesti näitlejatest, suur osa heliloojatest, paljud kunstnikud. Eesti elavikku on püütud mõtestada ligi 200 täispikas mängufilmis. Kroonika- (üle 8500 pala) ja dokumentaalfilmide (üle 2000) pildis on näha või kõnelevad neis kümned tuhanded inimesed, puudutatakse tuhandeid Eesti ajaloo- ja argielu sündmusi, näidatakse arvukalt paiku ja ehitisi, kõneldakse paljudest asutustest ja institutsioonidest. Võiks ju (tõsi küll – ideoloogiliste mõõndustega) öelda, et sajast Eesti filmiaastast vaatab meile vastu sada aastat Eesti elu.

Informatsioon Eesti filmide kohta on paljude aastate jooksul olnud hajutatud, lünklik ja kohati puudulik ning, nagu ka filmid ise, avalikkusele mitte just hõlpsasti kättesaadav. Vanemad filmograafiad trükiti piiratud mahus ja tiraažis. Nii neis kui ka viimastel aastatel loodud elektroonilistes veebiandmebaasides on filmide kirjeldamiseks kasutatud erinevaid tunnusteskeeme. Institutsionaalsed veebiandmebaasid (Filmiarhiiv, Tallinnfilm, Eesti Filmi Instituut) hõlmavad Eesti filmivaramu üksikuid osi, Filmiarhiivis on vaid 60% ringis kõigist saja aasta jooksul tehtud filmidest. Süsteemse infokogu vajadus, uued infotehnoloogilised võimalused ja Eesti filmi 100. aastapäev (2012) toetasid motivatsiooni Eesti filmivaramut tervikuna hõlmava andmebaasi loomiseks ja selle avalikkusele kättesaadavaks tegemiseks.

Töö algfaasi olulisim ülesanne oli luua Eesti filmi infosüsteemi (EFIS) alusmudel.

Teisisõnu, andmebaasi tegemisele tuli seada filmivaldkonnale omase kultuuri-ruumi (filmiteosed, Eesti filmilugu, filmielu, filmitegijad, filmiinstitutsioonid, filmiharidus) eripärast ja andmebaasi sihtrühmade oletatavatest vajadustest tulenevad selged, põhjendatud ja jõukohased eesmärgid. Need eesmärgid tuli seejärel realiseerida andmebaasi metaandmete struktuuris, filmitunnuste klassifitseerimise põhimõtetes, kirjete kompositsioonis, sisumärksõnastiku ja sellele vastava otsingusüsteemi ülesehituses. Metodoloogilisi lähtekehti pakkus erialakirjandus, eeskujuga võtta ka paljudes riikides tehtud nii eriotstarbelistest kui ka massitarbijaile mõeldud elektroonilistest filmiandmebaasidest, mida koostavad filmiarhiivid, filmielu korraldavad riigiasutused, filmitootjate–turundajate–levitajate huvi teenivad arendajad ja filmialaste meelelahutusportaalide loojad (vt vastavat ülevaadet www.efa.ee).

EFA peaeesmärk on Eesti filmikultuuri hõlmava põhjaliku andmekogu loomine laiale avalikkusele viisil, et „infolademe” süvakihid vastaksid ka valdkonna professionaalsetele (majanduslikele, arhiivialastele, filmiajaloolistele, filmiteaduslikele ja -hariduslikele) vajadustele. Praktikas tähendab see ulatusliku metaandmete süsteemi rakendamist filmitoote sisu, nende tegemises osalenud loovisikute ja asutuste kirjeldamiseks, lisaks mahukat uurimistööd filmograafilise, bibliograafilise, biograafilise, ajaloolise ja haridusliku andmestiku kogumiseks, fotokollektsioonide tegemist, filmitoote katkendite valimist ja filmide vaatamisvõimaluste tagamist veebis. Sellele toetuv detailne ja paljude valikutega filmiinfo otsing (*moving image information retrieval*) pea kõigi metaandmestikku sisestatud tunnuste põhjal võimaldab huvi- ja vajaduspõhise filmitoote leidmist, grupeerimist ja filtreerimist.

Olla üheaegselt kasulik paljudele sihtrühmadele on loomulikult ressursinõudlik eesmärk, mis tähendab filmitoote kirjeldamise mahu, keerukuse ja inimtöö vajaduse mitmekordset suurenemist lihtsa infoandmebaasi koostamisega võrreldes. Samas – Eesti mõõtu väikeriigi filmivaramu ja filmi-

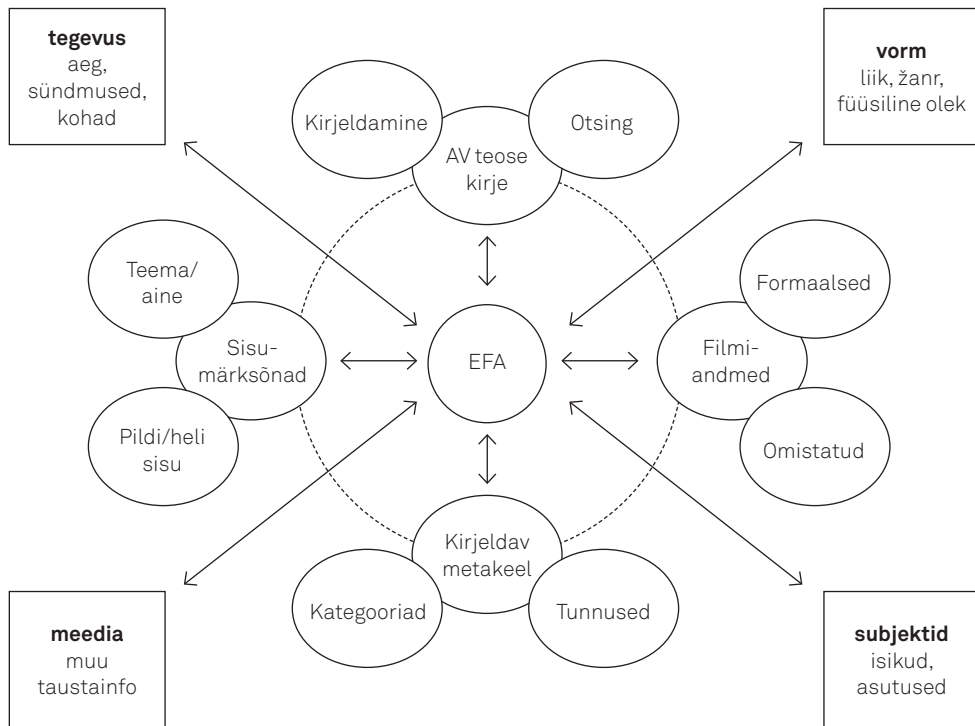
kultuuri puhul võiks detailsust, süvitiminekut ja põhjalikkust pidada mõistlikuks kuluks.

1.2. Metaandmestiku mudel

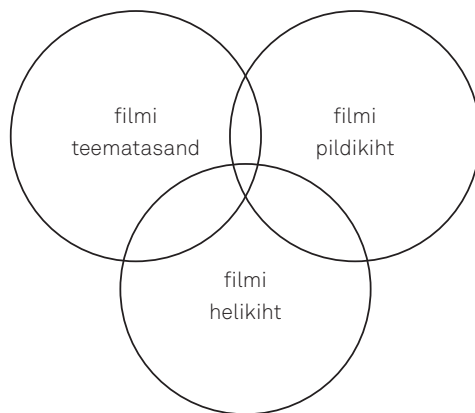
Filmitoote sisu ja vormi, samuti filmitegijate ja -asutuste metatunnuste üldine süsteemimudel on esitatud **joonisel 1**. Iga filmitoote, -tegijate ja -asutuste kirjeldamiseks võimalikult põhjalikult. Neid kirjeldavad metatunnused on liigitatud 14 kategooriasse: **pealkirjaala, vastutusala** (filmi peamised loojad, võttegrupi muud liikmed, filmi tootmisega seotud isikute täisnimekirja), **väljaannete ala** (filmi versioonid ja variatsioonid), **tootja- ja tootmisandmed** (tootjafirmad, kaastootjad, koostööpartnerid, tellijad ja sponsorid, lisaks ka mängufilmide sise- ja välisvõtete kohad), **levitaja ja levitandmed** (levitajad, esilinastused, linastumistulemused, festivalid ja auhinnad), **autoriõiguse andmed, teose ja materjali füüsiline kirjeldus** (teose erinevad kandjad ja tuletised filmil, videolindil või failina), teose ja materjali **liigipõhine klassifikatsioon**, lavastuslike teoste **žanriklassifikatsioon, rolliala** (mängufilmides esinevad näitlejad ja nende rollid), **sisukirjelduse ala** (sünopsis, sisukokkuvõte, muusikaleht, sisu märksõnastik, sealhulgas ka filmides nähtavad, rääkivad ja käsitletavat reaalsed isikud), **huviinfo** (filmi bibliograafia, arvustused, täiendav ja taustainfo, mälestused), **meediainfo** (publikatsioonid, visuaalid – fotod, filmilõigud, trailerid, helilõigud, ostujuht) ja **arhiiviandmed**. EFIS kodeerimisjuhendis kirjeldatud metaandmeid-üksiktunnuseid on erinevates kategooriates kokku ligi 350 (Šein 2009). Filmide sisu märksõnastik on omakorda jaotatud 13 sisukategooriaks ja need omakorda arvukateks alagrupideks (vt detailsemalt **joonis 6**).

Lisaks sisumärksõnadele on otsisõnades kõik filmikirjetes esinevad inimeste, kohtade, sündmuste ja institutsioonide nimed. Andmebaas võimaldab filmi-, sisu- ja isikuotsinguid nii üksiktunnuste (märksõnade) põhjal kui ka neid detailotsingus kombineerides. Lisaks saab otsingus kasutada mitmeid filtreid.

Igal filmil on andmebaasis oma **kirjellehekülg**, filmitoote põhiautoritel



JOONIS 1. Filmiandmebaasi metaandmestiku mudel.



JOONIS 2. Filmiteose tähenduskihid.

ja enamikul võttegruppide liikmetest, samuti kõigil filmitootmisega seotud institutsioonidel ja tootmisfirmadel on aga nn **persoonilehekülg**, mis annab ülevaate nende tegevusest ning sisaldab bio-graafilist ja filmograafilist andmestikku. Filmiandmebaasi veebiväljundis kuvatavad filmi, isiku ja asutuse kirjed on juba iseeneslikud märksõnastikud, sest kirjete olulisemad tunnused töötavad linkidena (vt www.efis.ee).

1.3. Teostus

Suure infokogu loomine ja andmestikuga täitmine on töömahukas, kulukas ja vajab partnereid. MTÜ Eesti Filmi Andmebaas initsiatiivil koondusid filmivaldkonnaga seotud võtmeorganisatsioonid – Kultuuriministeerium, Eesti Filmi Sihtasutus, Eesti Kultuurkapital, Eesti Rahvusringhääling, AS Tallinnfilm, Eesti Rahvusrhiiv, Eesti Kinoliit, Balti Filmi- ja Meediakool, Eesti Filmiajakirjanike Ühing ja Eesti Film 100 –, kelle esindajad sõlmisid 20. jaanuaril 2009 andmebaasi loomist toetava ühiste kavatsuste kokkuleppe.

Metaandmete süsteemi väljatöötamiseks kulus ligi kaks aastat. Selle põhjal õnnestus saada rahaline toetus Euroopa Regionaalarengu Fondi projektikonkursilt (rakendusüksus Riigi Infosüsteemide Arenduskeskus, RIA), korraldada riigihange Eesti Filmi Infosüsteemi (EFIS) arvutitarkvara ja sellele tugineva veebilahenduse loomiseks, mille 2011. aasta kesksuveys teostas seitsme firma seas riigihanke võitnud Mindware OÜ.

Andmebaasi kirjete (filmiteoste, isikute ja institutsioonide) suure arvu tõttu (üle 20 000) toimub rahvusfilmograafia koostamine etapikaupa kümne aasta jooksul (2008–2018). Andmebaasi esimene järk avati veebis 2012. aasta detsembris, Eesti filmi 100. aastapäeva tähistamisel. Siis said veebis (www.efis.ee) kättesaadavaks enamiku filmide **põhiandmed (osakirjed)**, 1912–1940 tehtud kroonika- ja dokumentaalfilmide (ligi 1100 teost) ja kõigi seni toodetud täispikkade mängufilmide **täiskirjed**, samuti valik Eesti tudengifilmide täiskirjetest perioodist 2005–2012. Andmebaasis

saab **täispikkuses vaadata** kõiki kuni 1940. aastani tehtud filme ning tele- ja tudengifilme, samuti treilereid või tutvustavaid lõike kõigist mängufilmidest. Täispikkuses nähtavate filmide arv sõltub eelkõige filmide digiteerimise tempost ja autoriõigustega seotud küsimuste lahendamisest.

Samm-sammult liidestatakse filmiandmebaas teiste Eesti filmi- ja mäluasutuste andmebaasidega, arvestatud on võimalusega sidestuda Euroopa Filmivärvaga (<http://www.europeanfilm-gateway.eu/>), mis vahendab ligipääsu Eesti digiteeritud filmivaramule internetiportaalil Europeana.

2. FILMIPILDI MÄRK-SÕNASTAMISE SPETSIIFIKA

Pildi traditsioonilisel märksõnastamisel (*image indexing*) transleeritakse üks meedium teise meediumi keelde, filmipilti indekseeritakse sõnade ja mõistete abil. Selle protsessi muudab keerukaks asjaolu, et ühe meediumi vahenditega väljendatud tuleb tähistada teise meediumi keeles (Menard 2010). Erinevalt pildi visuaalse sisu **automaatselt indekseerimisest** (vt. ka 3.3) nõuab pildi **kontseptuaalne indekseerimine** tekstilist (verbaalset) vaheastet ning see pole inimese vahenduseta teostatav, sest – vähesed valdkonnad välja arvatud – objektide ja nende tähenduse (eriti liikuva ja monteeritud pildi) masinidentifitseerimine on keerukas, kui mitte võimatu (Ménard 2009, Rasmussen 1997, Terris 2000). See on eriti ilmne abstraktsete mõistete puhul, mis tulenevad kujutatavate objektide või suhete tähendusest ja kontekstist – masinat on raske panna ära tundma mõisteid ja abstraktseid tähendusi, näiteks „elu pärast surma“, „ema-lapse suhted“, „kolonialism“, „homoseksuaalsus“, „aastatuhat“ jms.

Filmipildi verbaalset kirjeldamist sõnade abil on tekkivate piirangute tõttu võrreldud püüdega ehitada elavat puud saematerjalist. **Lahendus on siiski selles, et seda, mida saab pildis ära tunda, mõista, nimetada ja interpreteerida, saab ka märksõnastada.**

Filmiteose omapära on, et selle sisu- ja tähendusruum tekib sümbioosina

mitmetest „täendus- ja sisukihtidest” ja nende koosmõjust, see on filmipilti ja -helisse (teksti, dialoogi, muusikasse) erinevatel viisidel ja erinevate vahenditega kodeeritud. Filmikunsti spetsiifikast tuleneb omakorda, et valdavat osa sisust ja tähendusest püüavad filmiloojad reprodutseerida filmi pildireas, pilti „pannakse” enamasti see kontekstuaalne võti, mida vaataja vajab sisu tähenduse mõistmiseks.

Konkreetselt pildi- ja helirea ning filminarratiivi sisu üldistades saab sõnades teha üldisemaid sisumääratlusi filmi valdkonna, teema, aine, probleemi või sündmuse kohta ning normeeritud klassifikaatoritunnustele (näiteks UDK) vastavuse kohta. Nimetame neid deskriptoreid **filmi teematasandi** tunnusteks, mis moodustavad ühe filmi ainekiri kirjeldamise kihtidest (*headings*). Lisaks teematasandile saab eristada ja detailselt kirjeldada (märksõnastada) ka pildi- ja helirea sisu, need deskriptorid kirjeldavad siis vastavalt **filmi pildikihti ja helikihti**. Otse loomulik, et nimetatud tasandid ja tähenduskihid kattuvad ja suhestuvad (vt **joonis 2**) – filmi sisu ja tähendus kujuneb selles kolmetasandilises läbipõimunud ruumis. Filmis sisu ja tähenduse märksõnastamisel Eesti filmi andmebaasis käsitleme metodoloogiliselt pilti ja heli ühtse tervikliku „tekstina”.

Filmipildi sisu märksõnastamine (*subject indexing*) tähendab osundamist, kust vastavat sisu võib leida. Pildi sisuindeksi (aineloendi) loomisel kasutatakse üksik sõnu ja sõnaühendeid, mida nimetatakse märksõnadeks (ka deskriptoriteks, indeksiterminiteks). Märksõnad on üldisemat laadi „rajaleidjad”, märksõna on sisu (aine) „nimi”. Pildimärksõna tähistab „**nähtava põhist sisu**”, semiootiliselt väljendudes nii denotatiivset kui ka konnotatiivset sisu.

Pildiindeksid võivad tähistada kaadris nähtavaid asju, tegelasi ja tegevusi, kohti ja objekte (maja, vaher, kosk, kana, varemed, tikkimine, kaevamine); või olla sõnad, mis kirjeldavad episoodi, stseeni, pikemat lõiku või kogu filmi (külamiljöö, nõukogude aeg, kollektiviseerimine, vabadusvõitlus); või olla pigem assotsiatiivsed kui kirjeldavad (rahu-lik, jõulurahu, üleemeelik, ängistav, isamaa-

line, eestipärane). Kui just pole tegemist fakti- või indekseerimisveaga, ei saa enamasti väita, et ühed märksõnad on valed ja teised õiged, küll võivad teatud märksõnad olla täpsemad ja sobida paremini kui teised. Pildimärksõnade „semantiline haare” on lai, ulatudes kirjeldavatest-denotatiivsetest assotsiatiivsete-konnotatiivseteni. Valdavas enamuses on filmi märksõnad **pildimärksõnad**, need tähistavad nähtavat ja kaasnevat (tõlgenduslikku) sisu ja tähendusi, mis avanevad filmipildi kui „**visuaalse teksti**” kirjeldamise, analüüsi ja tõlgendamise kaudu.

Sageli kannab film tähenduslikku sisu (muuhulgas ka emotsioone ja üldistusi), mida pole otseselt filmi tegevustikus, ei tekstis ega helireas, ja mis tekib filmi vaatamise, tajumise, sellest arusaamise ja mõtestamise tulemusel. Filmis tähenduslik sisu sõltub sellest, kuidas vaataja filmi tajub, sellest aru saab, millele ta tähelepanu pöörab, kuidas ta filmi tähenduse enese jaoks „lahti võtab” ja „kokku pakib”, seda tunnetab ja väljendab.

Oluline indekseerimise aspekt on indekseerija „lähedus-kaugus” filmitoeses käsitletavast ajast, sündmustest, inimestest ja teemadest. 1930. või 1970. aastail toodetud filmide indekseerimine toimub ju „tagantjärele”, võrreldes näiteks 2010. aastate filmidega. Materjali modaalsus (suhe reaalsusega, Rafferty 2011) ja indekseerija teadmised ning kogemus mängivad olulist rolli märksõnade valikus.

Filmitoeses puhul tuleb silmas pidada, et sisu ja tähendust antakse edasi filmispetsiifiliste vahenditega, mis peituvad nii filmi struktuuris, selle elementides, filmi pildi- ja helireas (verbaalses tekstis ja muus heli/muusikalähenduses), samuti filmi väljendusvahendites (filmikeeles). Konkreetsetes filmikangas põimuvad kõik need lõimed loogiliselt, aga ka assotsiatiivsel moel terviklikuks tähendust kandvaks filmitekstiks. Filmis **struktuurielementidena** tuleb nimetada kaadrit, võtteplaani, stseeni, episoodi, aga ka filmi kui tervikteksti, filmikeele vahenditena **montaaživõtteid, plaani suurusi** (üld-, kesk-, suur, ülisuur, detail), **võttenurki-rakurse**, pildi liikumise kiirust

(kiirendamine, aeglustamine või stoppkader), värvi- ja valguslahendusi jm.

Filmipildi **kodeerimisühikuks** võib olla nii üksik filmikaader, võtteplaan (klipp, montaažiplaan), episood (osa stseenist, koha ühtsus), stseen (tegevuse ühtsus), lõik (kroonikas või dokumentaalfilmis, sisuline, temaatiline ühtsus) kui ka film tervikuna; igaüks neist võib kanda olulist sisu. Sama võtmesõnaga tähistatud, kuid kestuselt erineva pikkusega filmiühikutel võib olla sama mõistemaht. Kodeerimisühiku füüsilise maht ehk kestus märksõnastikus ei kajastu, see selgub andmebaasi kasutaja jaoks leitud lõigu läbivaatamisel. Kodeerimisühiku valimine põhineb igal konkreetset juhul andmebaasi toimetajate ühisel arusaamal pildi indekseerimise otstarbest. Protsessi subjektiivsust ei saa välistada, küll aga saab seda tasakaalustada toimetamis- ja kodeerimistöö põhimõtete kehtestamise ja nende järgimisega.

Filmi kirjeldamisel ja märksõnastamisel on pildi sisu sõnastamise (kodeerimise) põhimõtted ja võtted seega otsustava tähtsusega. Pildisisule nime andmine peaks olema süstemaatiline ja efektiivne – see peab aitama leida antud otsinguruumis sarnase pildisisuga filmitooteid ning pakkuma otsijale mõistlikku hulka objekte. Liiga kitsad või liiga ulatuslikud otsingutulemid on võrdselt ebasoovitavad, seetõttu on oluline lähtuda sobivatest ja asjakohastest funktsionaalsetest märksõnastamise põhimõtetest ja valdkonnapõhisest otstarbekusest vastavalt sõnastatud kultuuriväärtuskriteeriumitele (vt täpsemalt allpool).

3. FILMIPILDI MÄRKSÕNASTAMISE TEOREETILISED LÄHTEKOHAD

1.1. Panofsky kolm pildianalüüsi taset

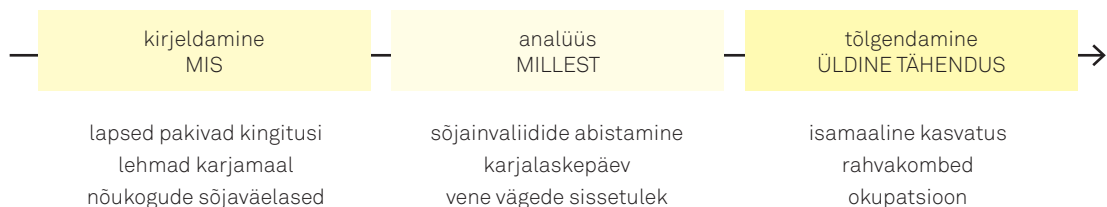
Paljud pildi märksõnastamise käsitlused toetuvad kunstiteoreetik Erwin Panofsky alustpanevale teorialele (1933) kolmest tasandist pildi sisu ja tähenduse avamisel, mida ta rakendas renessansi kunstiteoste analüüsimisel. Need on:

- **ikonograafiaelne**: pildi kirjeldamise üldine (*generic*) tase, mis vastab küsimusele, mis on pildil näha (*description of the image*),
- **ikonograafiline**: pildi konkreetse (*specific*) analüüsi tase, mis vastab küsimusele, millest pilt on, mida pilt sisuliselt kujutab (*analysis of what an image is about*),
- **ikonoloogiline**: pildi abstraktse (*abstract*) interpreteerimise tase, mis väidab, mida pilt tähendab (*interpretation of the meaning of an image*).

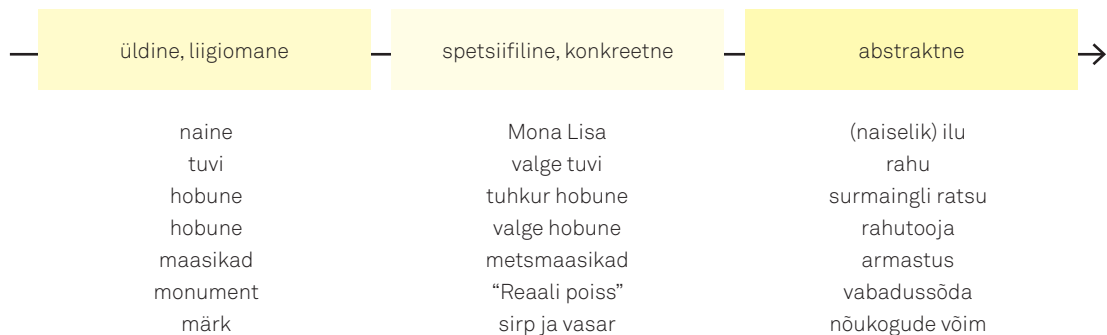
Panofsky skeem (vt selgitusi näiteks Neugebauer 2010, Rafferty 2011) osutab kolmele pildi sisu ja tähenduse mõistmises olulisele kognitiivsele protsessile. Lihtsustatult võiks öelda, et pildi (ja selle üksikute osade) puhul võime eristada seda, **mis on pildil näha** (mille pilt see on), **millest see pilt on** (mida see pilt konkreetset väljendab) ja mis on selle pildi **üldisem tähendus** või mõte. Skeemist tuletatud „**mis ja millest**” teooria (*Of and About Theory*, vt 3.2.) pani pildi sisu ja tähenduse uurimisele ja praktilisele märksõnastamisele fundamentaalse aluse. Skeemi kohandas infoteaduste tarvis üks tuntumaid pildi indekseerimise uurijaid Sara Shatford-Layne valdkonna ühes olulisemas artiklis “Analyzing the Subject of a Picture: A Theoretical Approach” (Shatford 1986), millele toetudes võiks filmipildi märksõnastamist kolmeastmelise protsessina (vt **joonis 3**) iseloomustada järgmiselt.

Kirjeldamine. Mis on pildil, mille pilt see on? Filmipildis kujutatud objektide ja tegevuste üldine kirjeldamine on reeglina **faktipõhine**, faktoloogiline (ütlemine „see on see”). Võib ka küsida, mis laadi see pilt on, millist meeoleolu see kannab? Pildi meeoleolu, emotsionaalse värvingu kirjeldus on **ekspressiivne**. Sellel märksõnastamise tasandil saab niisiis rakendada nii faktoloogilist kui ka ekspressiivset kirjeldamist.

Analüüs. Pilti analüüsidest saame väita, millist konkreetset sisulist tähendust pilt endas tegelikult kannab, millest see pilt on, mida see väljendab? (ütlemine „see



JOONIS 3. Näiteid kroonikafilmi pildisisu märksõnastamisest Eesti filmi andmebaasis.



JOONIS 4. Näiteid üldise, spetsiifilise ja abstraktse tasandi eristustest.

tähendab seda”). Analüüsi mõjutavad teadmised ja kontekst, pildi tähenduse mõistmine eeldab vastava kultuuri tundmist (ka laiema tausta, inimeste, olustiku, sündmuste, aja jms. tundmist ja äratundmist). Ka sellel tasandil on rakendatav nii faktoloogiline kui ka ekspressiivne lähene mine. Analüütiline märksõnastamine võib ka ebaõnnestuda: kui märksõnastaja ei „tunne asja ära”, ei saa sellest aru või ei tea selle tähendust, on oht, et osa filmi sisust või tähendusest jääb mõistmata, „läheb kaduma” ega kajastu sisuindeksis.

Tõlgendamine. Mis on pildi seesmine, sellele omane (*intrinsic*), ka olemuslik, üldisem, abstraktsem või sümboolne tähendus? Milliseid üldistusi see võimaldab, milliseid laiemaid tähendusi filmipilt loob? Kuna seda tähenduse taset on raske üheselt indekseerida, on infoteadlased enamasti seisukohal, et pildiindeks ei peaks sisaldama interpretatsioone (Shatford 1986, Hixson 2003, Neugebauer 2010). Artikli autori arvates on siiski just üldistamise ja tõlgendamise teel sõnastatud märksõnad väärtuslik märksõnastiku tasand, seda eriti väikeses, „hästi loetavas” kultuuriruumis. Muidugi, tulemuslik tõlgendamine nõuab kvalifitseeritud indekseerijate tööd.

3.2. Panofsky/Shatfordi mudel („mis”, „millest” ning üldine ja spetsiifiline „mis”)

Shatford (1986, 1994) modifitseeris Panofsky skeemis üldise, spetsiifilise ja abstraktse tasandi eristusi (vt **tabel 1**) ning osutas, et pildi sisu indekseerimisel tuleb ennekõike vastata kahele olulisele küsimusele: **mis** (*of, ofness*) ehk kes või mis on pildil, mille pilt see on? – see on ikooniline sisu (näiteks pildil on naerev inimene); ja **millest** (*about, aboutness*) ehk millest pilt kõneleb? – mida pilt sisuliselt väljendab ja tähendab, millist mõtet kannab ja sisaldab – see on ikonoloogiline ehk sümboolne sisu (näiteks naerva inimese pilt võib väljendada rõõmu, õnne, kahjurõõmu, hullumeelsust jms).

„Mis” (*of, ofness*) on samaaegselt nii **üldine** (*generic*) kui ka **spetsiifiline** (*specific*). Tuues eesti näite: kui näeme pildil silda, siis märksõna „sild” on üldine, kui me aga

tunneme ära, et see on Pirita jõe sild, siis saame spetsiifilise, konkreetse tähistuse „Pirita sild”. Peab kindlasti rõhutama, et pildi puhul saab sageli tähistada nii üldist kui ka spetsiifilist üheaegselt (Hixson 2003), kõik oleneb sellest, kas piirduakse vaid kirjeldusega või seatakse eesmärgiks ka analüüs ja tõlgendus (vt konkreetseid näiteid **joonis 4**).

„Millest” (*about, aboutness*) on pigem subjektiivne, abstraktne ja sümboolne. Sümboolne sisu kirjeldab emotsioone ja abstraktseid mõisteid (jällegi eesti näite abil: pildil on lehviv lipp, pilt võib aga sisult kanda tähendust Eesti, kodumaa, töotus, au ja väärikus jms).

Edasi pakkus Shatford igal tähenduse tasandil vaadelda nelja erinevat tahku (*facet*), mis vastavad küsimusele kes?, mida?, kus? ja millal? (vt täpsemalt 4.6.). Nii kujunes pildi tähenduse käsitlemise mudel, mida tuntakse 3 × 4 maatriksina ehk Panofsky/Shatfordi mudelina (JISC 2013), mida on kasutatud paljudes pildianalüüsi töödes ja praktilistes lahendustes, Eesti filmi andmebaasi sisumärksõnastiku ülesehituse põhimõtted kaasa arvatud.

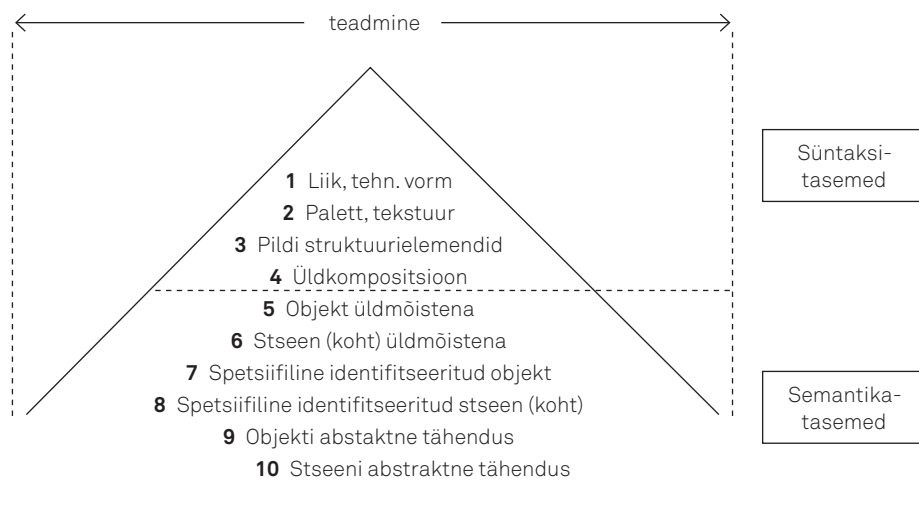
3.3. Automaatne ja manuaalne pildisisu indekseerimine

Mida rohkem audiovisuaalseid materjale ja teoseid digiteeritakse, mida enam neid digiarhiividesse ja veebiandmebaasidesse ladestub, seda suuremaks muutub vajadus suurte andmemassiivide kirjeldamise, nende indekseerimise (klassifitseerimise ja märksõnastamise) ning neil põhinevate otsingusüsteemide järele. Kuna tegemist on suurte andmemassiividega, siis on mõistatav, et viimasel aastakümnel on pildimaterjalide automaatse indekseerimise võimaluste vastu hakanud huvi tundma ka arvutiteadlased. CIBR (Content-Based Image Retrieval) valdkonnas ilmub palju uuringuid, milles kirjeldatakse arvutitehnoloogilisi võimalusi (algoritme) ära tunda ja märgistada seisva ja liikuva pildi **visuaalseid omadusi ja tunnuseid, visuaalset sisu** – värvilahendust, vormi, tekstuuri, kompositsiooni, eristada struktuurielemente, kaamera liikumist, montaaživõttestikku, analüüsida

TÄHENDUSE TASE	NÄITED	PANOFSKY KÄSITLUS	NOTES
Esimene	Faktipõhine: mees, naine, puu Ekspressiivne: lein, rahu, õnn, viha	Ikonoograafiaeelne (kirjeldus): esmane, loomulik aineestik, mõistmine eeldab vaid igapäevast kogemust Faktipõhine Ekspressiivne	Üldine objektide, sündmuste ja tegevuste kirjeldus. Faktipõhine: MIS (<i>of</i>), ekspressiivne: MILLEST (<i>about</i>)
Teine	Lootosasendis rõõmsameelse koguka mehe kujutis: MIS tasandil Buddha, MILLEST tasandil väljendab kaastunnet.	Ikonoograafiline (analüüs): teisene ehk konventsionaalne aineestik, mõistmine eeldab kultuurilaseid teadmisi	Spetsiifiline objektiivne MIS, müüdi, abstraksiooni või sümbolina MILLEST
Kolmas	Zanussi film <i>Illuminatsioon</i> : MIS tasandil näide <i>cinéma vérité</i> st ja MILLEST tasandil film teadusest, eetikast ja valgustusajastust. Hall kivimüür sümboliseerib sotsialismi (MILLEST)	Ikonoograafiline (interpretatsioon): olemuslik tähendus, mõistmine nõuab sünteesivõimet, teadmisi kunstist, sotsiaalsest ja kultuurilisest taustast	Ei ole mõeldud indekseerimiseks, erandina ehk vaid väga spetsialiseeritud valdkondades, kuna ühest kokkulepet tähenduse kohta on liiga keeruline saavutada

TABEL 1. Panofsky ja Shatfordi lähenemise võrdlus.

Allikas: Neugebauer, Thomas 2010.



JOONIS 5. Visuaalse sisu deskriptorite püramiid pildi indekseerimiseks.

Allikas: Jaimes, Chang 2000.

heli, otsida pildireast žanritunnuseid kandvaid formaalseid kombinatsioone jms. Infotehnoloogilises lähenemises eristatakse seetõttu visuaalsete elementide, võiks ehk öelda **kujutise märksõnastamist** (*content indexing*) ehk arvuti abil automaatselt pildi füüsiliste omaduste tuvastamist ja tähistamist **kontseptuaalsest märksõnastamisest** (*concept indexing, intellectual indexing*) ehk inimese poolt filmipildi tähenduste märksõnastamisest (vt ka Ménard 2007).

Sümbioosi neist kahest lähenemisest pakub Alejandro Jaimesi ja Shih-Fu Changi (2000) töö, mida autorid nimetavad visuaalse sisu deskriptorite püramiidiks, visuaalsete deskriptorite klassifikaatoriks (vt **joonis 5**). **Süntaksi tasemeid 1–4** ehk pildi visuaalset vormi, struktuuri, visuaalse objekti tajutavat füüsilist olekut (*directly related to percept*) saab tuvastada ja indekseerida automaatselt. Tasemed 5–10 ehk **semantilised tasemed** on seotud visuaalse tähendusega (*visual concept*): **üldine, spetsiifiline ja abstraktne objekti ja stseeni tase** (*generic, specific, and abstract object and scene*) kirjeldavad visuaalset sisu (*visual content*). Semantikatasandid toetuvad Panofsky/Shatfordi skeemile, stseeni mõiste sissetoomine sobib hästi filmipildi märksõnastamise lahendustes.

Masintuvastada saab tüübi- ja vormitasandit (*type, technique*), mis fikseerib sisuliselt kandja liigi (foto, maal, trükis, põhimõtteliselt siis ka film, video või fail), üldist värvipaletti ja tekstuuri (*global distribution*), spetsiifiliste pildielementide värvi ja tekstuuriomadusi (*local structure*), üldkompositsioon (*global composition*) kirjeldab pildielementide ja nende omaduste laotust ja asendit/korraldust pildil.

Objekti üldmõiste (*Generic Object*) näiteks on pildil nähtav: õun, lill, mees, tool, lehm, loom, sild jms. **Konkreetne identifitseeritud objekt** (*Specific Object*) fikseerib konkreetse objekti nime. Tähistamine põhineb teadmisel ja äratundmisel, on objektiivne, sest tunnus tugineb teadaolevatele faktidele isikutest või objektidest: näiteks Bill Clinton, Eiffeli torn, Russalka, Vabadussõja monument Tallinnas jms.

Objekti abstraksiooni tasandi (*Abstract Object*) määratlused tuginevad põhjalikumatele teadmistele või tõlgendusele, mida objekt tähistab, tähendab või sümboliseerib. Juba osutatud põhjustel on see indekseerimistasand vastuoluline. Kuigi enamasti üldtuntud tähendustega suuri probleeme ei teki, on tõlgendused siiski sageli subjektiivsed, kodeerijate ja kasutajate arvamused sama objekti tähendusest võivad olla väga erinevad, isegi vastandlikud. **Koha üldmõiste** (*Generic Scene*) näiteks oleks linn, maastik, siseruum, vaikelu, portree jms. **Konkreetne identifitseeritud koht** (*Specific Scene*) tugineb konkreetsele teadmisele ja teadaolevatele faktidele: näiteks Pariis, Vihterpalu, Times Square, Central Park, Kadriorg jms. **Abstraktne representatsiooni ja tähendustasand** (*Abstract Scene*) on pildi kui terviku tõlgendustasand, mis on samuti üsna subjektiivne. Terviku tähendust kirjeldatakse sageli emotsionaalsete või abstraktsete (atmosfäär, teema) terminitega: näiteks tervikstseeni abstraktset tähendust tähistavateks sõnadeks võivad olla kurbus, surm, võit, taevast või paradiis, kättemaks, kuritegu, võimu kehtestamine jms. Deskriptorite püramiidi kui metodoloogilise lähenemise sarnasus Panofsky/Shatfordi mudeliga on ilmne.

Käsitatud semantiliste tasandite eristus pildi sisu ja tähenduse indekseerimisel on inspireerinud ka Eesti filmi andmebaasi märksõnastamissüsteemi lahendust, samuti märksõnastiku struktuuri, kategooriate ja tunnuste liigitusi. Kuna Eestis ulatuslikum audiovisuaalteoste märksõnastamise praktika on alles tekkimas, tuleb märksõnastamise käigus teoreetilise skeemi kasutamise otstarbekust samm-sammult testida ning konkreetseid lahendusi vastavalt tekkivale kodeerimiskogemusele pidevalt täiendada ja arendada.

4. FILMITEOSTE MÄRK-SÕNASTAMISE PRAKTIKA

4.1. Märksõnastamise mõte

Filmipildi indekseerimisel on kaks eesmärki: (1) võimaldada leida teatud tunnustele vastavaid (sisu ja tähendusega) filme ning (2) luua võimalused filmide grupeerimiseks

teatud tunnuste põhjal (Shatford 1994).

Filmide märksõnastamine võimaldab suure hulga filmide seast **soovitavate otsimist (ja leidmist)** vastavalt otsijat huvitavatele formaalsetele ja sisulistele (meta)tunnustele, seda nii üksiktunnuste kui ka tunnuste gruppide põhiselt. Sisumärksõnastiku struktureerimiseks tuleb luua kasutajale mugav klassifikaator, püüdes seejuures aimata, mida filmidest enamasti otsitakse ja millised võiksid olla andmebaasi kasutavate erialagruppide spetsiifilised vajadused. Et Eesti filmivaramu ja filmiteosed kannavad endas olulist ajaloolist ja sotsiaalkultuurilist teavet, pilti Eesti elavikust, sündmustest ja inimestest, võib öelda, et märksõnaotsing peab pakuma ligipääsu just sellele sisu- ja tähendusruumile.

Indekseerimine üldisemalt võimaldab filme **liigitada (rühmitada)** ühe, mitme või paljude metatunnuste põhjal. Filmide formaalsete tunnuste alusel tehtud nimistutest on levinuimad loovisikute või filmifirmade **filmograafiad**. Sisumärksõnastamine võimaldab otsida (leida) teatud ainestiku, teema, sisu ja tähendusega filme ja neid vastavate tunnuste alusel rühmitada. Milleks on see vajalik? Esiteks on sarnaste tunnustega filmiteoste kõrvutamine ja võrdlemine sagedane uurimismeetod filmialases teadustöös, teiseks, teatud tunnuste alusel grupeeritud teosed võivad omada ühisjooni tunnuste põhjal, mis selguvad alles saadud grupi lähemal analüüsimisel ja võimaldavad seega uut teadmist, ja kolmandaks, leitud teoste grupist saab valida konkreetsele kasutamise eesmärgile kõige sobivama filminäite või vaste (vt ka Shatford 1994). Seega on peamiseks teaduslikud, õppimisega seotud ja praktilised vajadused. Kindlasti on grupeerimine abiks ka huvipõhise valiku tegemiseks meelelahutuslikel eesmärkidel.

Mida siis enamasti filmiandmebaasidest otsitakse? Turnerile (1997) toetudes on olulisemad pildihuvi lähtekohad **võttesuund või plaani suurus** (suur plaan, üldplaan, panoraam), **värv või domineeriv emotsioon** (voogav kuldne viljapõld), **aeg või ajaperiood** (aastaaeg, kümnendid: näiteks 1950. aas-

tate miljö, nõukogude aeg), **geograafiline koht** (linn, paik, loodusvorm, 1960. aastate Tallinna tänavapilt). Kindlasti on otsingu sihikuks ka **teema või valdkond** (töö, protsess, suhted, tegevus, amet), **fakt või sündmus**, teatud **subjekt või objekt** (näitlejad, filmitegijad, inimesed filmis, asutused, esemed, ehitused), teatud **sümboolset tähendus kandev pilt, liigi- ja žanriotsingud** (ka filmi tonaalsus, muud professionaalsed väljendusvõtted). Hea märksõnastik püüab olla kooskõlas valdava otsihuviga.

Kindlasti vajavad andmebaasi sügavama filmihuviga professionaalid: filmitegijad, filmitootjad ja levitajad, filmiuurijad, filmist kirjutavad ajakirjanikud ja saateid tegevad teleajakirjanikud/produtsendid, filmiüliõpilased, samuti õpetajad, kes võivad olla huvitatud illustreeriva ja mõtestava lisamaterjali leidmisest ja audiovisuaalsest toest läbivate teemade käsitlemisel erinevates ainetundides (Joassoone, Peterson 2011). Samas on film atraktiivne ja huvitav kultuuri- ja meelelahutusvaldkond, mistõttu võib eeldada ka laiemat publikuhuvi. Seetõttu on mõistlik teha metakirjeldus üheaegselt nii **professionaalse kasutajakonna** kui ka **laiema auditooriumi info- ja meelelahutuslikke vajadusi** silmas pidades, neid ühendavaks lüliks oleks filmiteoste sisu kultuurilis-ajalooline kaalukus.

Andmebaasi tarbijaid silmas pidades on märksõnastiku põhjalikkuse ja kasutatavuse mõttes oluline teadmine, et erinevad kasutajarühmad on huvitatud erinevast sisust, igal kasutajarühmal on erinevatel põhjustel ja ajal erinevad sisuhuvid, sama sisu võib erinevatele inimestele tähendada erinevaid asju ning omada erineval ajal erinevat tähendust ka sama inimese jaoks.

4.2. Märksõnastiku lahendused

Filmidele **infoligipääsu sisendite** (*access points*) indekseerimisel on kaks peamist võtet, mille alusel tegevindekseerijad jagunevad enamasti kahe erineva põhimõtte pooldajateks: esimese leeri selge eelistus on olemasolevad **normeeritud märksõnastikud** (*controlled vocabulary*, Eesti juhtumil näiteks Eesti Märksõnastik, EMS) ja sellel baseeruv **ohjatud märksõnastamine** (Nilbe

1994), teiste valik on **vaba märksõnastamine** märksõnastamise praktikas tekkiva sõnastiku (*uncontrolled vocabulary*) põhjal. Esimesel juhul võetakse märksõnad fikseeritud tesaurustest (*thesauri*), ühtlustatud klassifikatsiooniskeemidest (*classification schema*) või aineleenditest (*lists of subject headings*), teisel juhul on märksõnadena kasutatavad kõik loomuliku keele vahendid. Viimastel aastatel on lisandunud ka kolmas lahendus, kus andmebaaside kasutajad lisavad (*tagging*) piltidele märksõnu ise (*collaborative tagging, social tagging, user-based indexing*). Tugevaid külgi, aga ka piiranguid ja puudusi on kõigil neil indekseerimisviisidel (vt näiteks Ménard 2007; Besiki, Jørgensen, Wu 2012).

Praktikas (ka meie filmiandmebaasi praktikas) püüavad andmebaasi tegijad konkreetset sisumärksõnastamise protsessis moodustunud esialgset vaba märksõnade kogumit klassifitseerida ja korrastada ning vajaduse korral ühtlustada ka normeeritud märksõnastikega. Sobiva lahendi leidmiseks eelneb kontrollitud (ka normeeritud) märksõnastikele sageli vaba märksõnastamise staadium: Eesti filmi andmebaas alustas vaba märksõnastamisega sisumärksõnade, üksikasjade ja filmi teema kategoorias, samas filmide formaalsete andmete, UDK, geograafiliste deskriptorite ja isikunimede puhul toetuti olemasolevatele normeeritud sõnastikele või eelnevalt korrastatud liigitustele ja loeteludele.

Kuigi filmiandmebaasi sisumärksõnastamisel ei rakendatud baassõnastikku, kehtestati siiski vaba märksõnastamist „ohjavad” reeglid, mis fikseeriti andmebaasi kodeerimisraamatus (Šein 2009). Sama eesmärki teenis ka tekkiva märksõnastiku klassifitseerimine kategooriatesse ja alamgruppidesse. See võimaldas ja suunas indekseerijaid vajaduse korral kasutama nii ülem-, baas- kui ka alamkategooriaid (mööbliese – tool – kiiktool; linnukasvatus – kanafarm – kanad), samuti ainsust, mitmust ja üldsõna (tüdruk – tüdrukud – lapsed) ning „nihutas” märksõnastikku sageli oluliste üldistuste poole. Tuleb tõdeda, et kategoriseerimisküsimuste

lahendamine nii eel- kui ka järelkoordinatsiooni võtete abil on vältimatult üks indekseerimisprotsessi keerukamaid komponente, samuti mõjutab grupitunnuste määratlus otseselt infootsingu tulemuslikkust (Rorissa, Iyer 2008).

Märksõnastike, klassifikaatorite, tunnuste kategooriate, alamgruppide ja rühmitamisvõtete kasutamisel kerkib sageli probleem sellest, kui efektiivselt suudavad tavakasutajad reaalses infootsingus kasutada andmebaaside loojate poolt konstrueeritud tunnusteskeeme. Mitmetes uurimustes (Herzum 2003, Bron *et al.* 2011), on näidatud, et professionaalide loodud andmeskeemid ei toeta kuigi efektiivselt tavakasutajate otsinguvajadusi, sest tavakasutajal on keerukas „läbi näha” andmeskeemi või siis ilmnevad erinevused mõistete kasutamisel ja otsinguküsimuste formuleerimises. Üks võimalus seda probleemi lahendada on märksõnastiku ja selle struktuuri võimalikult detailne avamine kasutajatele, mida oleme Eesti filmi andmebaasi puhul ka teinud.

4.3. Filmiteoste sisu märksõnastamise vormilised aspektid

Iga filmiteose märksõnastik esitatakse andmebaasis nimistuna, **märksõnade hulk** konkreetse filmiteose kirjeldamiseks ei ole piiratud, kuid peaks olema otstarbekas ja relevantne. Praktika näitab, et kroonikafilmi märksõnastik võiks sisaldada 10–30 ja mängufilmide märksõnastik 50–60 märksõna.

Märksõnad on otstarbekas esitada **tähestikulises järjekorras**. Kuigi bibliograafilistes märksõnastikes kasutatakse enamasti mitmust, leiame, et filmi märksõnastikus on otstarbekas eristada **mitmust ja ainsust**, mis tuleneb audiovisuaalse kujutise märksõnastamise eripärast (näiteks hobuse ja hobuste, vanuri ja vanurite otsing annaks erinevaid filmitulemusi). Märksõnadeks võivad olla **üksik sõna**, nimisõnad, omadussõnad, tegusõnad, mäarsõnad, arvsõnad, samuti objektide, isikute, sündmuste, kohtade, institutsioonide nimed. Üksikuid märksõnu peab saama otsingus kombineerida, samas võib

märksõna ise olla eelkoordineeritud kombinatsioon **kahest või enamast sõnast**, nii moodustuvad **märksõnastringid**.

Filmide märksõnastamisel tekkinud märksõnade kogumid moodustavad klassifitseeritud ja grupeeritud märksõnaloendeid, mida saab sirvida esinemise sageduse, tähestiku ja valdkonna/teemapõhiselt. Märksõnaloenditest (filmiliikide, žanrite, filmitegijate, -näitlejate, filmides näidatavate konkreetsete isikute, kohanimede, sündmuste, asutuste ja institutsioonide nimistutest) moodustub universaalne filmiteste sisusõnastik. Eesti filmi andmebaasi konkreetse märksõnastiku struktuur, sisuja vormikategooriate ja tunnuste loendid ning klassifikaatorid on alates 2012. aasta lõpust kättesaadavad andmebaasi veebis (www.efis.ee). **Filmiandmebaasi märksõnastik on otsijale andmebaasis avatud ja kättesaadav**, mis hõlbustab oluliselt otsinguprotsessi.

4.4. Filmiteseid kirjeldavad metatunnused

Nagu eespool öeldud, võib andmebaasi mõistes kõiki filmide sisu, vormi ja füüsilisi tunnuseid, samuti filmi loonud isikute ja institutsioonide nimesid käsitleda märksõnadena (vt **joonis 1**). Võime tõdeda, et märksõnastatud filmiteseid sisaldav andmebaas töötab peamiselt ja ennekõike märksõnapõhise filmiteseid kirjeldava ja otsiva masinana. Filmiteseid kirjeldava metasüsteemi tunnused on märksõnad, mis töötavad samaaegselt otsingusõnadena.

Shatford eristab nelja oluliste pilditunnuste kategooriat: **biograafilised tunnused, sisutunnused, liigi- ja füüsilised tunnused ning kaasnevad tunnused** (*biographical, subject, exemplified and relationship attributes*). Biograafilised tunnused on enamasti objektiivsed: teose loojate nimi, loomise aeg, teose pealkiri, asukoht, hind, variandid. Sisu kirjeldavad tunnused ehk „mis” ja „millest” (nii üldises kui ka spetsiifilises mõttes) on fikseeritavad neljas kategoorias: (1) aeg, (2) ruum (koht), (3) tegevused ja sündmused ning (4) objektid ja subjektid. Filme tehakse ja neid levitatakse erinevatel kandjatel, filmiga kaasnevad objektid on näiteks

stsenaarium, kriitilised artiklid, uurimused, huvitekstid, veebilehed, auhinnad, fotod filmi saamisloost, filmiplakatid, mälestused jms. Kõigi nende tunnuste põhjal saab filme kirjeldada, indekseerida (märksõnastada) ja muuta andmebaasis otsitavaks.

Märksõnu võib liigitada ka **formaalseteks andmeteks** (tunnusteks), mis on objektiivsed ja vaieldamatud: info, mis sisaldub nt pealkirjaalas (originaal-, paralleel- ja tööpealkirjad), vastutusallas (filmi põhiloojad), tootja- ja tootmisandmetes; leviandmed, filmitese füüsilised parameetrid, osatäitjad mängufilmides, tootmisgruppide täielik isikufilm, geograafilised kohad, filmis aset leidvad konkreetset (ajaloo-) sündmused, filmides näidatavad konkreetset isikud, asutused ja institutsioonid jms ning kokkuleppelisteks märksõnadeks ehk **omistatud andmeteks** (tunnusteks): näiteks liigi- ja žanritunnused, ainevaldkonna ja teema määratlused, abstraktsioonid ja interpretatsioonid. Filmitese spetsiifilist aine- ja pildisisu kirjeldavaid märksõnu nimetame **sisumärksõnadeks**.

Huvitav erijuhtum on pildis mitte-nähtava (sündmuste, kohtade, objektide) märksõnastamine. Eesti filmi andmebaasis oleme neid nimetanud **seotud märksõnadeks**, mis osutavad filmist puudevatele, kuid ainepõhiselt filmiga seotud objektidele või asjaoludele. Näiteks intervjuufilmis Eesti olümplastega 1936 ei näe me kaadreid Berliini olümpiamängudest, ometi on selle märksõna lisamine oluline neile, kes otsivad olümpiamängudega seotud materjale. Nii näiteks on Inglismaa filmi- ja teleaariivis 20 filmi märksõnaga Titanic, kuid vaid ühes filmis on näha laeva ennast. Ometi on kõigil materjalidel oluline suhe Titanicu looga (Terris 2000).

Lisaks formaalsetele ja omistatud andmetele saab koguda ka **kaasnevaid filmiandmeid**, mis tekivad filmide linastamisel, levitamisel, müügil, säilitamisel ja kriitilisel käsitlusel erinevates publikatsioonides (juhtlause, sünopsis, sisukirjeldus, festivaliinfo ja auhinnad, erinevate väljaannete ja versioonide kirjeldus, levitajad ja müüjad, autoriõiguste omanikud, filmi kohta ilmunud kirjutiste bibliograafia

(artiklid, raamatud, veebileheküljed), huviinfo, mälestused, arhiivinfo jm.

Praktikas on kasutusel palju kataloogimise skeeme (Sam 2009). Märksõnastamise võtete seas võiks eristada universaalset, **üldist märksõnastamist**, süsteemset **valdkonnaspetsiifilist süvamärksõnastamist** ja otstarbepõhist **funktsionaalset märksõnastamist**.

Vastavalt sellele kujunevad andmebaasides välja üldised, ontoloogilised, tüpoloogilised, funktsionaalsed, ka väärtuspõhised märksõnastikud. Üks huvitavamaid valdkonnaspetsiifilisi sisumärksõnastikke, mis on loodud andmebaasi toimetajate ja filmihuviliste koostöös, on Internet Movie Database (IMDb), milles on üle 5000 märksõna (<http://www.imdb.com/Sections/Keywords/>). Tüpoloogilise märksõnastamise näitena võiks tuua Filmigenoomi (Movie Genome) projekti andmebaasi (<http://www.jinni.com>), milles filmid on märksõnastatud ja tüpologiseeritavad ekspertide (ja andmebaasi kasutajate) poolt koostatud ja filmidele omistatud koondtunnuste (filmi laad, süžee, žanr, aeg, koht, sihtauditoorium ja auhinna) põhjal. Heaks näiteks audiovisuaalteoste valdkonnaspetsiifilisest märksõnastamisest on vähemus- ja põlisrahvaste kultuuripärandit käsitleva ontoloogilise kirjelduse loomine, mida seejärel kasutati indekseerimissõnastikuna (Stockinger 2007; ka veebiportaal <http://www.culturalheritage.fr/EN/>). Eesti filmi andmebaasi üldmärksõnastikus oli vaba märksõnastamise faasis 2013. aasta keskpaigas 13 märksõnagrupi (vt tabel 2) umbes 40 000 märksõna, neist sisumärksõnu umbes 15 000, filmides näidatavate isikute nimesid (v.a näitlejad) oli umbes 8500.

5. MÄRKSÕNASTAMISE PRAKTILISI KÜSIMUSI

5.1. Vastuolu kirjeldamis- ja otsingueesmärkide vahel

Olgu tegemist kroonikapala, dokumentaal-, mängu- või animafilmiga – kui filmiteost käsitleda mitmetasandilise audiovisuaalse tekstina, ilmnevad sisu ja tähenduste indekseerimisprotsessis teatud piirangud. Infootsingu vaatepunktist on otstarbekas

küsida, kas konkreetsete märksõnade kasutamine on oluline, kas just neid on vaja? Kuna konkreetsele otsingutunnusele vastavaid objekte võib olla tuhandeid, huvitab ratsionaalset infootsijat relevantne ja kvaliteetne, otsingu põhjusele või eesmärgile vastav, „kasutata“ otsitulemus. Kui konkreetne aine, huvipakkuv objekt või subjekt ei ole antud filmis otsinguküsimusele just kõige täpsem (või tehniliselt mitte kõige parem) vaste, siis oleks ehk õigem seda mitte tähistada ning eeldada, et sama sisu on kindlasti paremini esitatud mõnes muus filmis. Näiteks näeme merd sadades eesti filmides – kui infootsija tahab leida häid emotsionaalselt laetud kaadreid ja stseene mere erinevatest olekutest ja seisunditest, on sajad otsinguvasted märksõnale „meri“ tajutavad pigem infomürana. Teisisõnu, me võime teha filmile täiusliku sisukirjelduse, märksõnastada detailselt ja põhjalikult kogu nähtava ainekogu ja sisu, aga samas võib selline märksõnapilt otsingu aspektis „toota“ ebapraktilist teavet.

Sisumärksõnastamise põhjalikkuse ja detailsusega on otseselt seotud ka grupeerimise võimalused. Piltlikult on seda mugav selgitada „kuudil istuva koera“ kaasusega. Esimene võimalus on jääda märksõnastamisel piisavalt üldiseks: indekseerida vaid koerad ja kuudid, otsija leiab siis kõik filmid, milles esinevad nii koerad kui ka kuudid, vaatab need läbi ja ehk leiab mõnest filmist ka kuudil istuva koera. Teine võimalus on minna väga spetsiifiliseks: indekseerida nii istuvad, konti närvivad, urisevad, haukuvad kui ka magavad koerad ja, juhul kui kaadris on ka koerakuut, märkida ka, kas koer teeb neid tegevusi kuudi peal, kuudi juures või kuudis. Sama probleemi illustreerib näide, milliseid märksõnu pakkusid indekseerijad andmebaasi koolituskursusel hobuse märksõnastamiseks Sulev Keeduse „Georgicas“ (vt tabel 3).

Märksõnastamise oluline praktiline küsimus on ka „kas tervik või osa?“. Üldjuhul – kas märksõna „käsi“ asendab edukalt märksõna „sõrmed“? Millistest indekseerimispõhimõtetest juhinduda ja milliseid rakendada, jääb reeglina iga konkreetse andmebaasi koostajate otsustada.

SISUMÄRKSONAD	ÜKSIKASJAD	KOHANIMED
Film–TV	Ained ja materjalid	Asumid
Filmi tonaalsus	Ehitised ja rajatised	Pinnavormid
Inim-, pere- ja paarisuhted	Esemed, asjad, väikevahendid	Saared, poolsaared, laiud
Keha, elu ja tervis	Kunstiesemed	Teed, tänavad, väljakud, pargid
Kultuur ja kunstid	Laevad ja muud veesõidukid	Veekogud, jõed, kosed, rannad
Loodus, linnud, kalad	Liiklusvahendid	Välismaa
Loomu- ja isikuomadused	Masinad, seadmed, relvad	Ülevaated, panoraamid
Miljöö, kombed, aeg	Riietusesemed	EHITISED JA OBJEKTID
Riik, poliitika, ajalugu	Sümboolika, lipud, vapid, märgid	Ausambad ja monumendid
Seksuuaalsuhted	Toiduained, söök ja jook	Hooned
Sport		Jaamad, lennuväljad, sadamad
Taimed		Kalmistud, hauad
Tegelased	ASUTUSED JA INSTITUTSIOONID	Kindlused, kloostrid, mõisad
Tegevused	Kaitseväge ja Kaitseliidu üksused	Kirikud
Tunded, emotsioonid	Kauplused	Spordirajatised
Tegevuskohad	Koolid, ülikoolid, haridusasutused	Sillad
Tegevusvaldkonnad	Kunstiasutused	Turud
	Lasteasutused	ISIKUD FILMIS
TEEMAD (MÄNGUFILM)	Seltsid, liidud, ühingud	Poliitikud ja riigitegelased
12 gruppi, vt www.efis.ee	Meditsiin ja tervishoid	Teadlased
	Muusikakollektiivid	Sportlased
TEEMAD (KROONIKA)	Põllumajandusasutused	Kultuuritegelased, kultuuritöötajad
12 gruppi, vt www.efis.ee	Riigi- ja valitsusasutused	Tööstustöölised ja -juhid
	Sportdivõistkonnad	Põllumajandustöötajad ja -juhid
AINEVALDKOND UDK	Teadusasutused	PIIRDAATUMID
ligi 300, vt www.efis.ee	Teenindusasutused	Kuupäev või vahemik
	Tööstusettevõtted	Aasta
	Välismaa asutused	Kümnend

TABEL 2. Eesti filmi andmebaasi sisumärksõnastiku liigitus.

hobune Misjonär, haavatud hobune, hirmutatud hobune, hullunud hobune, perutav hobune, valge hobune, hobuse piinlemine, hobune silotornis, hobusega heinte tallamine, hobusega sõitmine, hobuse jooksutamine, hobuse taltsutamine, hobuse treenimine, hobuse rangid seinal, surev hobune, hobuse surm

TABEL 3. Hobusega seotud võimalikud märksõnad „Georgicas” (16 märksõna).

Indekseerijad jagunevad kahte leeri ka küsimuses kas „märksõnastada kõik, mida tead ja näed” või „märksõnastada kindla peale, seega tähistada vaid seda, mis kindlasti ei muutu”. Viimasel juhul tuleks seega vältida kahtlusi, ebakindlust, oletusi, andmeid, mis võivad edaspidi täpsustada (Reser 2012). Sõltuvalt andmebaasi loovast institutsioonist ja andmebaasi spetsiifikast lahendatakse need küsimused alati isemoodi, kaalukeeleks akadeemilised, funktsionaalsed ja kasutajapõhised kriteeriumid.

Niisiis läheb filmiteose sisu põhjaliku kirjeldamise taotlus sageli vastuollu otsingu ratsionaalsuse nõudega. Andmebaas ei tohiks petta otsijate ootusi tühise või ebatäpse otsingutulemusega. See paneb suure vastutuse toimetajale/andmesisestajale, kes peavad kaaluma, kas antud sisu ja pilt on konkreetse märksõna ja tähenduse suhtes piisavalt informatiivne, täpne ja milline võiks olla selle „kasutusväärtus”. On abiks, kui andmesisestaja või toimetaja küsib endalt, et kui ta ise peaks otsima (ütlemes teletimetajana, kooliõpetajana või filmiuurijana) konkreetseid filmiepisoode ja filmikaadreid antud teema või mõiste kohta, kas ta siis vajaks otsingutulemustes just seda filmiteost?

Seega suunab **pragmaatiline märksõnastamine** tähistama vaid olulist ja iseloomulikku, millel on tajutav kasutamise- ja informatiivne/kultuuriline väärtus. Sisulise kirjeldamise täielikkuse huvidest lähtuv positsioon aga eeldab tingimata kogu nähtava sisu, pildi ja heli tähendusruumi „äraindekseerimist”, vahel ka vastutuse võtmist võimalike ebatäpsuste pärast (mida saab muidugi alati parandada). Kuna rangeid normatiivseid reegleid märksõnastamiseks on raske anda, määravad otstarbeka tasakaalupunkti ennekõike andmebaasi loomise põhiväärtused ning kasutamispädevus ja kasutajatelt tulev tagasiside.

5.2. Filmiteose sisu

märksõnastamise kriteeriumid

Kuigi filmiteoste pildi märksõnastamine on enamasti väga nüansirikas, saab siiski sõnastada mitmeid ideaalseid kriteeriume ja printsiipe:

- **objektikohasus:** liigitamise ja märksõnastamise süsteem ja protsess on objektiga kooskõlas, st, sobib filmide märksõnastamiseks;
- **asjakohasus ja sobivus:** märksõna tähistab ja aitab leida relevantset, märksõna sisule/tähendusele/mahule vastavat sisu ja objekte, märksõna mõistemaht „katab” objekti hästi, valitud sõna on informatiivne;
- **kohane põhjalikkus ja kaalukus:** filmiteose märksõnade kogum katab olulise osa filmi sisu- ja tähendusruumist, märksõnastikus ei lähe filmi tähendusruumist midagi olulist kaduma;
- **otstarbekus:** märksõnaotsing pakub mõistlikku hulka objekte, välditakse infomüra ja inforisustust, infootsija ei jää „tühjade kätega”.

5.3. Metodoloogilisi näpunäited filmipildi märksõnastajale

Konkreetse märksõna valikul on seega ikka ja jälle kohane küsida, kas see on oluline? Kas see, kuidas antud teema, aine, sündmus, protsess, objekt või subjekt on antud filmis esitatud, on relevantne ja kvaliteetne vaste antud märksõnale ja kas seda on mõtet indekseerida? Kogenud filmibibliograafid peavad mõistlikuks filmide märksõnastamise käigus esitada järgmisi küsimusi (kohandatud, arendatud ja täiendatud Shatfordi ja Hixsoni järgi):

- **Kes?** Kes või mis on kaadris? On need inimesed, lapsed, mehed, naised, lehmad, hobused, puud, ehitised? Kas kaadris on konkreetseid inimesi, kelle nime te teate, asjad, kohad või sündmused, mida te oskate nimetada konkreetse nimega?
- **Mida?** Mida need olendid või objektid filmipildis teevad? Milline on nende seisund või olek? Milliseid emotsioone see seisund või olek võiks üldjuhul vaatajas esile

SISUMÄRKSONADE KLASSIFIKATSIOON	VÕIMALIKUD SISUMÄRKSONAD
Miljö, atmosfäär, stiil	sotsialistlik paatos, nõukogude ajastu vaim, poeetilisus, elu saarel, võõrandumine maast, külaelu, kaos, intiimsus, traagiline lõpplahendus
Emotsioonid, tunded, suhted	rõõm, mure, kannatus, ülevus, hirm, perekonnasuhted, kaastunne, armastustseen, seksistseen, flirt, sõprus, kaklus, üksildus, võistlus, truudusetus, valitsemine, vägivald, üllatus, viha, vanamehe ja poisi suhted, purjusolek, sallimatus, sentimentaalsus
Aeg, ajastu	sotsialismiaeg, misjonitöö, noorusaeg, lapsepõlv, vanadus
Elu kulg, inimese seisund	noorus, vanadus, vaesus, kurbus, valu, kannatus, suremine, paanika, haigus, enesehävivate, emadus, ajaga võidujooks, tervenemine
Tegevused	ravitsemine, helistamine, pommitamine, vaatluste tegemine, palvetamine, tõlkimine, kalapüük, paadiga sõitmine, ratsutamine, silo tegemine, mee söömine, kala puhastamine, palvetamine, nõelamine, rääkima õpetamine, tervenemine, lugemine, koraali laulmine, lõõtsa tallamine, suremine
Tegelased	tumm poiss, vanamees, ema, inimesed
Ruum, koht	meri, saar, kirik, orel, kantsel, vahitorn, lennuk, rong, paat, roostes laev, auto, haigla, jalgratas, mesitaru, taevas, polügoon, angaar, voodi, tee, silo, maja
Abstraktne mõiste, üldistus, mõju, tagajärg	poeetilisus, laastatud küla, põlvkondade side, feminism, nostalgia, aatelisus, üksmeel, ajastu vaim, koloniaalmisjonitöö, võõrandumine maast, lunastuse võimalikkus
Loodus, loodusnähtused, aastaajad	torm, raju, vihm, lained, päikesepaiste, pilved, äike, udu, päikesetõus
Taimed, loomad, putukad	hobune, mesilane, kala, kajakas, silo, kõrkjad

TABEL 4. „Georgica” sisumärksõnade pilv.

- kutsuda või millist tähendust edastada? Milliseid valdkondi või abstraktseid ideid need tegevused või olud sümboliseerivad?
- **Kus?** Kus on filmipilt „üles võetud”? Milline on selle koha ruumiline määratlus? Kas see on identifitseeritav kindla geograafilise kohana? On see kunstlikult loodud ruum? Kas ruumiline määratlus – koht, paik, mida pilt näitab või milles tegevus toimub – on märksõnastamise mõttes oluline?
 - **Millal?** Millal võib olla nii konkreetne kuupäev, aasta, aasta-aeg kui ka teatud periood, ajavahemik, kümnend. Kas ajamääratlus on oluline? Kas te suudate selle määratleda? Kas ajal võib olla ka mingi üldisem sõnaline määratlus?

Nende sisutasandite märksõnastamisel võiks püüda rakendada veel mitmeid olulisi metodoloogilisi printsiipe ja praktilisi juhendeid:

- märksõnastage pilti, mis on ennekõike oluline filmiloo jutustamise, narratiivi, tähenduse ja mõtte edasiandmise seisukohast (tähistage olulised objektid, suhted, tähendused, sümbolid jms.);
- pidage silmas märksõna „kaugust” loo sõlmest (sõlmedest), peateljest, milleks on kasulik välja joonistada filmiteose sisu märksõnapuu (vt konkreetne näide **joonis 6**), mida „kaugemas harus” konkreetne märksõna on, seda küsitavam on selle kasutamine;
- pöörake tähelepanu spetsiifilistele, unikaalsetele ja huvitavatele detailidele, sest need võivad osutada oluliseks;
- pöörake tähelepanu montaažile, sest just montaaživõtted on filmispetsiifiline instrument tähenduse edasiandmiseks;
- kui vaja, indekseerige eraldi ka heli ja graafikat;

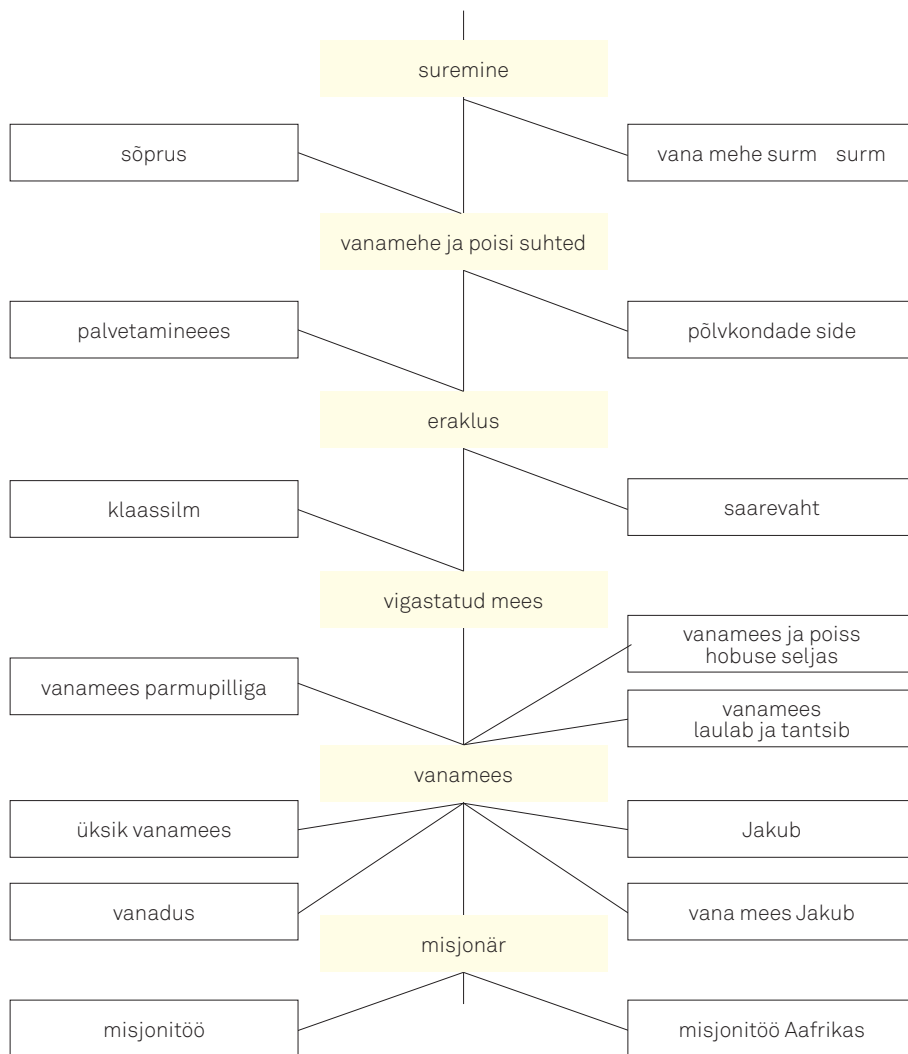
- püüdke konkreetse kõrval tabada ja tähistada ka üldisemat: teema valdkondlikku kuuluvust, väärtuspõhisust, kultuurilist või sotsiaalset, laiemalt funktsionaalset kaalukust, emotsioone ja tonaalsust, abstraktset tähendust;
- jälgige žanripetsiifikat ja autori mõtte konteksti;
- kasutage võimaluse korral ka filmispetsiifilisi märksõnu, need võivad olla abiks filmiprofessionaalidele;
- püüdke „märksõnade pilve” kokkupaneku järel selgusele jõuda, kuivõrd see vastab filmi üldisele sõnumile või tähendusele, suruge see seejärel kokku võimalikult otstarbeka lõpptulemuseni.

Kogu kirjeldatud protsessi üks probleemsemaid aspekte on selle paratamatu subjektiivsus. Toon siinkohal näite Sulev Keeduse „Georgica” märksõnapilvest ja selle klassifitseerimise võimalustest märksõnagruppideks. **Tabelis 4** on toodud andmebaasi sisutoimetajate/kodeerijate õppetööna loodud sisumärksõnade kogum ja selle võimalik klassifikatsioon märksõnagruppideks, **joonisel 6** aga filmi peategelase Jakubi kirjelduse „märksõnapuu” (joonise vasakus tulpas on kümne sisestaja pakutud Jakubit tähistavad märksõnad).

Tabelis on 128 sisumärksõna filmi kohta ja 34 märksõna filmi ühe peategelase kohta. Otsustamises selle üle, kas kõigi nende kasutamine „Georgica” märksõnastikus on otstarbekohane, väljendubki kogu märksõnastamisprotsessi keerukas olemus.

KOKKUVÕTTEKS

Eesti elektroonilise rahvusfilmograafia loomine on paljude võlgade tasumine ja valgete laikude kustutamine. Andmebaasi eesmärk on leevendada „infonälga” filmivaldas ja aidata kaasa filmivaldkonna arengule. Riigi arhiivides on vaid umbes 2/3 kõigist Eestis toodetud filmidest. Eestis tuleks luua filmiregister. Tuleb soodustada filmiloo põhjalikumaid uuringuid. Koostamata



- | | |
|--------------------------|---------------------------------|
| eraklus (4) | sõprus (3) |
| Jakub (7) | vana mehe surm (4) |
| klaassilm (4) | vanadus |
| misjonitöö | vanamees ja poiss hobuse seljas |
| misjonär (15) | vanamees Jakub |
| vana mees (7) | vanamees parmupilliga |
| misjonärina Aafrikas (2) | vanamees tantsib ja laulab |
| palvetamine (4) | vanamehe ja poisi suhted (7) |
| põlvkondade side (3) | vigastatud mees |
| saarevaht (4) | üksik vanamees |
| suremine (3) | |

JOONIS 6. Jakubi märksõnapuu „Georgicas” („lähedus-kauguskriteerium”).

on filmitegijate biograafilise leksikon. Korrastamist vajab filmistatistika. Seni enamasti killustatud ja erinevate andmeskeemide alusel kogutud filmitaave peab muutuma süsteemseks, usaldusväärseks ja järjepidevaks.

Laiemas plaanis teenib andmebaas eesmärki vääristada Eesti filmipärandit ja edendada seeläbi teadmisi, haridus- ja vaimuelu. Filmivaramu pakub mitmekesiseid võimalusi kasutamiseks koolis: aine- ja keeleõpetuse avardamiseks, õppekavade läbivate teemade käsitlemiseks, filmiõppeks ja filmi kasutamiseks loovõppes. Ligipääs filmivaramule peaks soodustama filmide kasutamist uusloomes, digiõpikute koostamist, pakkuma materjali teadustööks erinevates valdkondades ja arvutitehnoloogiliste meetodite laiendamiseks humanitaarteadustes (*digital humanities*).

Artikli ilmumise hetkel oli andmebaas avalikus kasutuses olnud 10 kuud ja kogunud selle ajaga 132 000 külastust, neist 70% unikaalseid kasutajaid, 10% kasutajatest on välismaalt – ligi 25 riigist kõikidelt mandritelt. Selle aja jooksul on vaadatud 470 000 lehekülge. Seega huvi on.

Paralleelselt eestikeelsega valmib ka andmebaasi ingliskeelne versioon, mis avardab Eesti filmivälja veelgi, üle veebi ja rahvusvaheliselt, igaühele sobival ekraanil, ajal ja paigas. Lähemate aastate aktiivsus peaks meile laiema avama nii Euroopa Filmivärava (European Film Gateway, www.europeanfilmgateway.eu) kui ka Europeana veebientsüklopeedia (www.europeana.eu) võimalused. Nagu öeldud, vaatab meile filmidest vastu meie elatud aja ja inimeste lugu. See muudab filmide kirjeldamise, indekseerimise (märksõnastamise) ja otsingusüsteemide arenduse oluliseks ligipääsuvõtmeks kogu filmikorpusele ja sellest tulenevalt ka filmipõhisele kultuuri-, teadus- ja vaimuelule.

KASUTATUD KIRJANDUS

- Besiki, S.; Jørgensen, C.; Wu, S.** 2012. 'Establishing the value of socially-created metadata to image indexing'. – *Library & Information Science Research*, 34, 99–109.
- Bron, M.; Nack, F.; van Gorp, J.; de Rijke, M.** 2011. 'Exploratory search in an audio-visual archive: Evaluating a professional search tool for non-professional users'. – *Proceedings of the 1st European Workshop on Human-Computer Interaction and Information Retrieval (EuroHCIR 2011)*, Newcastle, UK, 4 July.
- Enser, P.** 2008. 'The evolution of visual information retrieval'. – *Journal of Information Science*, 34, 4, 531–546.
- Herzum, M.** 2003. 'Requests for information from a film archive: a case study of multimedia retrieval'. – *Journal of Documentation*, 59, 2, 168–186.
- JISC Digital Media 2013.** Approaches to Describing Images. <http://www.jiscdigitalmedia.ac.uk/guide/approaches-to-describing-images> (1 June 2013).
- Joossoone, D.; Peterson, E.** 2011. *Video ja televisioon lõimitud keele- ja aineõppe toetajana*. Tartu: AS Atlex.
- Hixson, C.** 2003. *Analyzing the Subject of a Picture: Guidelines*. University of Oregon Libraries, http://library.uoregon.edu/sites/default/files/data/diglib/meta/Image_Subj_Analysis.pdf (28 September 2013).
- Internet Movie Database (IMDb) sisu märksõnastik.** <http://www.imdb.com/Sections/Keywords/> (1 June 2013).
- Jaimes, A.; Chang, S.-F.** 2000. 'A conceptual framework for indexing visual information at multiple levels'. – *IS&T/SPIE Internet Imaging*, 3964. http://www.ee.columbia.edu/ln/dvmm/publications/00/ajaimes-spie00_internet.pdf (1 June 2013).
- Nilbe, S.** 1994. *Liigitamine ja märksõnastamine arvutikataloogis*. Tartu Ülikooli raamatukogu. <http://www.utlib.ee/ee/konverentsid/1994/black.html> (1 June 2013).
- Ménard, E.** 2007. 'Image indexing: How can I find a nice pair of Italian shoes?' – *Bulletin of the American Society of Information Science and Technology*, October/November, 22–25.
- Ménard, E.** 2009. 'Images: Indexing for accessibility in a multi-lingual environment – challenges and perspectives'. – *The Indexer*, 27, 2, 70–76.
- Ménard, E.** 2010. 'Image indexing and retrieval: Challenges and new perspectives'. – *The Canadian Journal of Information and Library Science*, 34, 3, 245–248.
- Neugebauer, T.** 2010. 'Image indexing'. – *The Indexer*, 28, 3, 98–103.
- Rafferty, P.** 2011. 'Informative tagging of images: The importance of modality in interpretation'. – *Knowledge Organization*, 38, 4, 283–298.
- Rasmussen, E. M.** 1997. 'Indexing images'. – *Annual Review of Information Science and Technology*, 32, 169–196.
- Reser, G.** 2012. 'What not to embed: Is it better not to embed certain cultural heritage metadata in images?' – *Visual Resources Association Bulletin*, 39, 1, 5.
- Rorissa, A.; Iyer, H.** 2008. 'Theories of cognition and image categorization: What category labels reveal about basic level theory'. – *Journal of the American Society of Information Science and Technology*, 59, 9, 1383–1392.
- Shatford, S.** 1986. 'Analyzing the subject of a picture: A theoretical approach'. – *Cataloging & Classification Quarterly*, 6, 3, 39–62.
- Shatford Layne, S.** 1994. 'Some issues in the indexing of images'. – *Journal of the American Society of Information Science*, 45, 583–588.
- Stewart, B.** 2010. 'Getting the picture: An exploratory study of current indexing practices in providing subject access to historic photographs'. – *The Canadian Journal of Information and Library Science (La Revue canadienne des sciences de l'information et de bibliothéconomie)*, 34, 3, 297–327.
- Svenonius, E.** 1994. 'Access to nonbook materials: The limits of subject indexing for visual and aural languages'. – *Journal of the American Society for Information Science*, 45, 8, 600–606.
- Stockinger, P.** 2007. 'The PCI ontology: Presentation and discussion of an ontology for the description and indexing of audiovisual resources in the field of cultural heritage of minorities and indigenous people'. – *Proceeding of Specific Targeted Research Project IST Call 4 – FP6-2004-IST-4*. LOGOS: Knowledge-on-Demand for Ubiquitous Learning.
- Terris, O.** 2000. 'What you don't see and don't hear': Subject indexing moving images'. – *Multimedia Information and Technology*, 26, 1, 57–60.
- The Movie Genome.** <http://www.jinni.com/movie-genome.html> (1 July 2013).
- Turner, J. M.** 1997. 'Indexing pictures: some considerations'. – *Annual Meeting Council on Botanical and Horticultural Libraries, Jardin botanique de Montréal*. <http://turner.ebsi.umontreal.ca/english/texts/cbh197.htm> (1 June 2013).
- Sam, H.** 2009. 'Cataloging and preservation of moving images: A survey of organisations and initiatives'. – *PNLA Quarterly*, 73, 4.
- Šein, H.** 2012. *Filmiandmebaaside annoteeritud kogu*. Tallinn: Eesti Filmi Andmebaas. <http://www.efa.ee> (27 September 2013).

Summary

2012 was the year of film in Estonia, when the 100th anniversary of Estonian film was celebrated. One of the most significant undertakings planned for this occasion was the creation of the Estonian film database (electronic national filmography). Performing this large-scale task was undertaken by the NPO Estonian Film Database, launched in 2007. The main objective of the undertaking was to form a complete Estonian national filmography within ten years (2007–2018) and make it available in a web environment to everyone interested, both in Estonia and abroad. The access to the database was opened in late fall, 2012 (www.efis.ee).

Together with newsreels, the number of produced items reaches over 12 000. Feature films, documentaries and popular films, anima, television, educational programmes, advertising films and newsreels form a rich collection of the life, history, culture and people of Estonia. Nearly 3 000 filmmakers and most Estonian actors and actresses have participated in creating the Estonian film heritage. Several thousand people, events, places, buildings, offices and institutions in Estonia participate in or are mentioned in the films. In addition, the films are addressing several thousand people shown or talking in films. The electronic database opens the film treasury in a summarised way, employing a variety of possibilities offered by modern electronic databases.

A metadata system and coding instructions were prepared for each film, person and institution in the extensive space of attributes with search options, which combines the interactive features of a film directory and bibliographical, biographical databases. Each film is described

as thoroughly as possible. The attributes of films contain data about the subject, genre, authors, cast, production team, locations, producers, copyrights and distributors of films and about the technical parameters of films, as well as the bibliography of films, references to the reviews, articles, books published about films and the makers of films, digitised frames and pictures from films, trailers and promotional clips, scripts, memories of the makers and other interesting details. The subject content of films is indexed in 12 categories and related sub-groups and enables the search of films by plot/subject content, physical items, themes of newsreels and feature films, people, time, events, locations, building sites and institutions. In addition, films are indexed by a film-adapted UDC. As a result, more than 50 000 keywords enable thorough multi-layered content and subject search. All filmmakers are given their personal websites, which provides an overview of their creative careers and filmographies.

The electronic film database is interfaced with other similar databases at the Estonian Public Broadcasting, film archive of the National Archives, National Library and the Baltic Film and Media School of the Tallinn University. The web interface offers the possibility to enter with an ID-card and allows advance into several digital storages, where it is possible to view the films produced and purchase them for streaming. The filmography is interfaced with social networks (Facebook, Twitter) and is aiming the possibility to interlink it with the European Film Gateway in the future, thus offering access to a digitised film treasury through Europeana. The database is aimed at film professionals, teachers, students, researchers and the general public as the target audience.

Among others, the key issues of cultural databases draw on the approaches and solutions for information retrieval and are relying in particular on the principles of conceptual (intellectual) subject indexing of audiovisual artefacts. Inspired by classical works of Panofsky, Shatford, Turner and others regarding image description,

analysis and interpretation the article covers some main issues regarding options for a multifunctional film indexing metadata. The text tackles different aspects of the description of moving images for public needs in general and also describes the specific details of the system, developed for deep keywording of Estonian films. The rationale, limits and disputable issues as well as our experience and basic suggestions for professional indexers who are undertaking these kind of tasks are also revealed.