

Aufgabenstellung für eine Abschlussarbeit

# Automatische Generierung von Texten aus Multimedia Feature Graphen

Eine Erweiterung des GMAF Frameworks.

## Lehrgebiet Multimedia und Internetanwendungen

Prof. Dr.-Ing. Matthias Hemmje

### Fernuniversität Hagen

Universitätsstrasse 1, D-58097 Hagen

Raum: 1 F09, Gebäude 3 (Informatikzentrum, Universitätsstr. 1)

Tel.: +49 2331 987-304

Fax: +49 2331 987-4487

Internet: <http://www.fernuni-hagen.de>

**Bearbeitungszeitraum:** ab März 2022

**Betreuer:** Dipl. Inform. (Univ.) Stefan Wagenpfeil

E-Mail: [stefan.wagenpfeil@fernuni-hagen.de](mailto:stefan.wagenpfeil@fernuni-hagen.de)

Skype: stefan\_wagenpfeil

## Lehrgebiet Multimedia und Internetanwendungen:

- Daten- und Dokumentenmanagement
- Informations- und Wissensmanagement
- Multimedia
- Informationssysteme und Datenbanken
- Informationsvisualisierung
- Information Retrieval
- Visuelle Mensch-Maschine-Interaktion
- Content- und Wissensmanagement im Internet
- Semantic Web
- Digitale Langzeitarchivierung
- Virtuelle Forschungsumgebungen
- Big Data Analyse
- Analyse natürlicher Sprache
- Berufliche Weiterbildung und E-Learning
- Industrie 4.0 und "Factories of the Future"

1 **Abstract:** Multimedia Feature Graphen sind Datenstrukturen, in denen Multimedia Features  
2 verwaltet werden können. Derartige Features entstehen typischerweise durch Analyseverfahren,  
3 wie z.B. Objekterkennung. Das Generic Multimedia Analysis Framework enthält mehrere solcher  
4 Analyseverfahren und verfügt über Algorithmen, um die jeweiligen Multimedia Features zu  
5 erfassen. Erklärbare, semantische Multimedia Feature Graphen (ESMMFG) repräsentieren diese  
6 Graphen in Form von Phrasenstruktur-Bäumen, die die Basis für die Erzeugung von korrekten,  
7 natursprachlichen Sätzen darstellen. Aufgrund der hohen Zahl von Knoten und Kanten in  
8 einem Multimedia Graphen, werden diese Sätze jedoch sehr lang und komplex. Daher wird ein  
9 Sprachmodell entwickelt, welches durch Transformationen einen einfacheren Phrasenstruktur-  
10 Baum erstellt und somit bessere semantische Erklärungen für Multimedia Feature Graphen  
11 ermöglicht.

12 **Keywords:** GMAF, MMFG, Graph Algorithmen, Multimedia Features, Feature Fusion

13 Die nachfolgenden Beschreibungen für die Abschlussarbeit im Bereich Informatik  
14 benennen die nötigen Aufgaben und Teilaufgaben und sollen ebenfalls als Übersicht  
15 und Leitfaden für die zu erbringenden Leistungen dienen. Zur Bearbeitung der nachfol-  
16 genden Aufgaben werden folgende Fähigkeiten und Kenntnisse vorausgesetzt:

- 17 • **Wissenschaftliches Arbeiten:** strukturiertes, analytisches Vorgehen auf Basis der  
18 gültigen Konventionen für Aufbau, Zitierweise und inhaltliche Bearbeitung von  
19 forschungsrelevanten Themen sind die Voraussetzung für das Schreiben einer  
20 Bachelor- oder Master-These. Hierzu gehören die eigenständige Literaturrecherche,  
21 die selbständige Aneignung von Wissen und die Umsetzung von theoretischen  
22 Konzepten in die Praxis.
- 23 • **Programmiersprachen / Beschreibungssprachen:** Java, XML. Von Vorteil sind Ken-  
24 ntnisse im Umgang mit Eclipse und ein grundlegendes Verständnis von Ontologien  
25 und RDFs.
- 26 • **Sonstige Systeme oder Konzepte:** Aufbau und Handhabung von Graph-Strukturen,  
27 Objektorientierung, grundlegende Konzepte der semantischen Modellierung, RDF,  
28 Semantic Web.

## 29 1. Themenbeschreibung

30 Das Generic Multimedia Analysis Framework (GMAF) ist eine Entwicklung am  
 31 Lehrgebiet für Multimedia und Internetanwendungen. Es ist in der Lage, verschiedene  
 32 Multimedia-Datentypen (z.B. Bild, Text, Video) zu verarbeiten und diverse Algorithmen  
 33 zur Feature-Extraktion auf die entsprechenden Dateien anzuwenden. Die zentrale  
 34 Datenstruktur im GMAF ist der Multimedia Feature Graph (MMFG), in dem somit  
 35 sämtliche Features eines (oder mehrerer) Multimedia Objekte verwaltet werden. MMFGs  
 36 sind gerichtete Graphen, die tausende Knoten und Kanten beinhalten können. Eine  
 37 Visualisierung ist in Figure ?? gegeben.

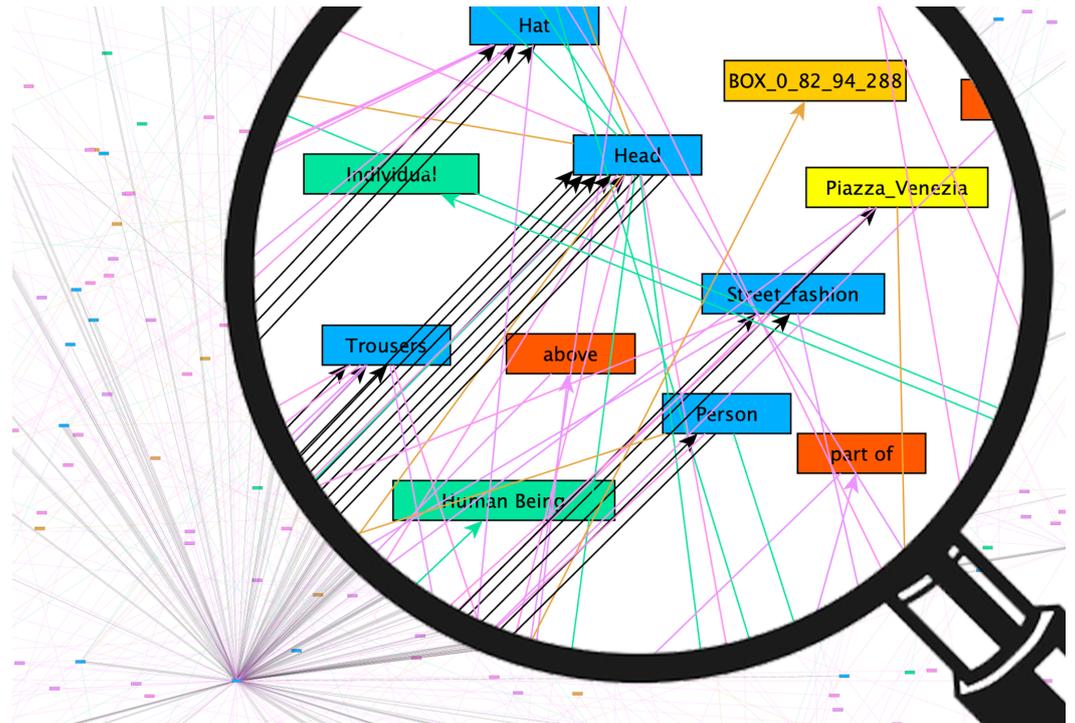


Figure 1. MMFG Visualisierung (Ausschnitt).

38 Im Rahmen der bestehenden Forschung am Lehrgebiet wurden diese MMFGs  
 39 bereits in semantische (SMMFG) und erklärbare (ESMMFG) Datenstrukturen überführt  
 40 (siehe Figur 2). Hierbei werden Konzepte aus der Grammatik- und Sprachtheorie  
 41 angewandt, um aus einer syntaktischen MMFG Struktur eine semantische und erklärbare  
 42 Phrasenstruktur zu erzeugen. Ein Phrasenstruktur-Baum ist ein Konzept aus  
 43 der Grammatiktheorie, und beinhaltet die Grundbausteine für die Konstruktion von  
 44 natursprachlichen Sätzen (z.B. Nominalphrasen, Verbalphrasen, Präpositionalphrasen,  
 45 usw.).

46 Da in einem MMFG (und den zugehörigen SMMFGs und ESMMFGs) jedoch eine  
 47 Vielzahl von Knoten und Kanten (und somit Informationen) gespeichert ist, werden  
 48 automatisch generierte Texte typischerweise sehr lang und kompliziert zu lesen - was im  
 49 Widerspruch zum Ziel der "Erklärbarkeit" steht. Im Rahmen dieser Abschlussarbeit wird  
 50 ein Sprachmodell entwickelt, welches in bestehende Strukturen des GMAF eingebettet  
 51 werden kann. Dieses Sprachmodell verfügt über diverse Einstellmöglichkeiten, um die  
 52 Länge und auch die semantische Tiefe des zu erzeugenden Textes zu beeinflussen und  
 53 erhält einen sehr umfangreichen Phrasenstruktur-Baum als Eingabeparameter, welcher  
 54 dann in einen menschenverständliche Sätze transformiert wird. Ein sog. *LanguageModel*  
 55 ist im GMAF bereits für diesen Fall vorgesehen.

56 Im Rahmen dieser Abschlussarbeit setzen Sie sich mit bestehenden Konzepten,  
 57 Tools, APIs, und Produkten für die Erzeugung von natürlicher Sprache auseinander  
 58 und implementieren ein (oder mehrere) Sprachmodelle für das GMAF auf Basis von

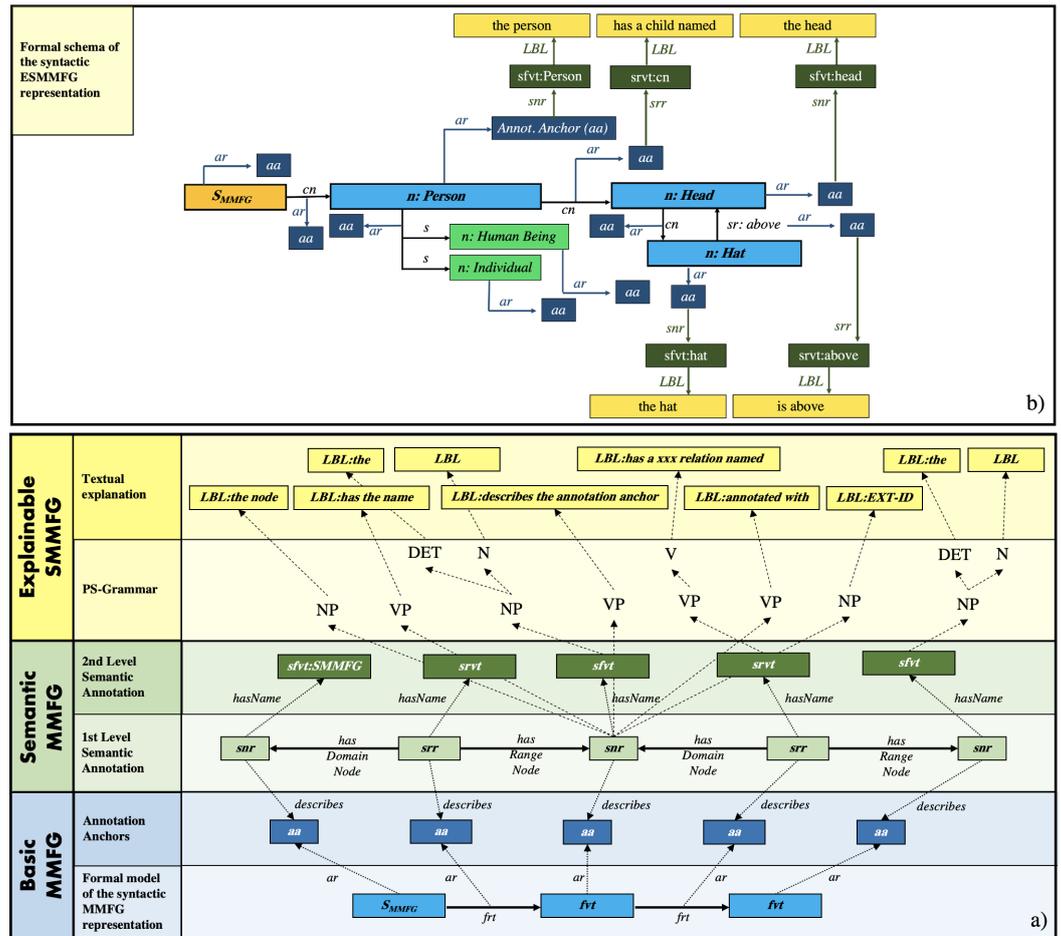


Figure 2. Erklärbare ESMMFGs und der zugehörige Phrasenstruktur-Baum (PS-Tree).

59 Phrasenstruktur-Bäumen. Im einfachsten Fall können Sie bestehende APIs nutzen, im  
 60 kompliziertesten Fall erzeugen Sie Transformationen auf dem Phrasenstruktur-Baum,  
 61 um die passende Tiefe von Formulierung und Semantik zu erreichen. Durch die Ein-  
 62 stellmöglichkeiten ist Ihre Lösung in der Lage, z.B. auch Abkürzungen oder Zusammen-  
 63 fassungen von Texten oder den Beschreibungen von Multimedia-Objekten zu erzeugen.  
 64 Eine Anbindung einer Text-To-Speech API kann ebenfalls Bestandteil Ihrer Arbeit sein,  
 65 um dem Benutzer der Anwendung einen zusätzlichen Mehrwert zu bieten.

66 **2. Theoretischer Teil**

67 Der theoretische Teil der Arbeit wird auf den Konzepten von Draper (User centered  
 68 system design) [29] und Nunamaker [30] aufgebaut und methodisch hergeleitet. Als  
 69 Modellierung wird die UML [31] verwendet. Dies bedeutet, dass sich auch die Doku-  
 70 mentstruktur von Exposé und Abschlussarbeit auf die Nunamaker Methodik stützt  
 71 und entsprechende Forschungsziele der Nunamaker-Typen (Beobachtung, Theoriebil-  
 72 dung, Implementierung und Experiment) definiert werden. Im theoretischen Teil der  
 73 Arbeit formulieren Sie zunächst Ihre Forschungsfragen und leiten die zugehörigen  
 74 Forschungsziele ab. Anschließend stellen Sie den Stand der Wissenschaft und Technik  
 75 vor und führen die für das Thema relevanten Konzepte ein. Auf dieser Basis kann dann  
 76 die Modellierung stattfinden.

77 **3. Praktischer Teil**

78 Der praktische Teil der Arbeit enthält die programmiertechnische Umsetzung der  
 79 Modellierung und die experimentelle Überprüfung. Als Programmiersprache kommt  
 80 typischerweise Java zum Einsatz. Die Verwendung von Standard-Entwicklungstools

81 wie Eclipse, GIT, Maven, Javadoc wird vorausgesetzt. Im Rahmen das praktischen Teils  
82 erfolgt auch die Erstellung von Test-Cases und deren Dokumentation, eine Evaluierung  
83 der Ergebnisse inklusive der Wahl und Begründung einer geeigneten Evaluierungs-  
84 Methodik, sowie ein Ausblick auf zukünftige Ergänzungen. Im Rahmen der Imple-  
85 mentierung folgen Sie den GoF Design Patterns und orientieren Ihre Arbeit an den  
86 bestehenden Strukturen. Hierbei sind insbesondere das *Command Pattern*, das *Abstract*  
87 *Factory Pattern*, *Singleton Strukturen* und das *Model-View-Controller Prinzip* zu berück-  
88 sichtigen. Den Einsatz und die Verwendung der Patterns stimmen Sie mit Ihrem Betreuer  
89 ab.

#### 90 4. Betreuung

91 Die inhaltliche und fachliche Betreuung der Arbeit wird von **Stefan Wagenpfeil**  
92 übernommen. Hierzu steht Ihnen jedereit die Kommunikation per Skype, Telefon,  
93 E-Mail zur Verfügung, wöchentliche Jour-Fixes mit dem Betreuer helfen Ihnen, die  
94 Arbeitsergebnisse strukturiert voranzutreiben und geben Ihnen die Gelegenheit, durch  
95 unmittelbares Feedback zügig und effizient die Ziele der Arbeit zu erreichen. Diverse  
96 Quality-Gates stellen sicher, dass die Thesis auch die Anforderungen am Lehrstuhl  
97 erfüllt und somit zu einem erfolgreichen Ergebnis führen wird.

#### 98 5. Aufgaben im Vorfeld

99 Bevor Sie sich dafür entscheiden, das Thema zu bearbeiten, sollten Sie sich einen  
100 groben Überblick zu den benötigten Voraussetzungen oder Hintergründen im Rahmen  
101 einer Recherche verschaffen. Im Literaturverzeichnis dieses Dokuments sind eine Reihe  
102 von Quellen hierzu aufgeführt. Anschließend beginnen Sie mit der Erstellung eines  
103 Exposé, welches den genauen Rahmen Ihrer Arbeit absteckt. Die Gliederung des Exposé  
104 sieht hierbei wie folgt aus:

##### 105 1. Einleitung (Motivation der Arbeit)

106 1.1. **Forschungsthema** (Beschreibung des Themas, Einordnung in ein übergeordnetes  
107 Themengebiet, Relevanz der Arbeit)

108 1.2. **Problembeschreibung** (Formale Beschreibung des Problems durch die Definition  
109 von Problemfeldern)

110 1.3. **Forschungsfragen** (Definition der Forschungsfragen zu den Problemfeldern)

111 1.4. **Methodik und Ableitung der Forschungsziele** (inklusive Nunamaker-Methodik)

112 1.5. **Ansatz** (Beschreibung, wie die Forschungsziele gruppiert werden und wo sie im  
113 Rahmen der späteren Arbeit behandelt werden)

114 1.6. **Aufbau** (Festlegung der Struktur der Arbeit)

115 1.7. **Zeitplan** (inklusive Meilensteinplanung und Gantt-Diagramm auf Basis der  
116 Forschungsziele)

##### 117 1.8. Zusammenfassung

118 Die Ergebnisse fassen Sie im Rahmen eines Dokumentes (Exposé) in ca. 10 bis 15  
119 Seiten zusammen. Das Exposé ist die Basis für die spätere Abschlussarbeit und gibt  
120 einen für alle Beteiligten definierten Rahmen der Arbeit vor. Dadurch werden unter  
121 Anderem folgende Punkte sicher gestellt:

- 122 • Überprüfung der formalen und schriftlichen Fähigkeiten des Studenten zur Erstel-  
123 lung und Ausarbeitung einer Abschlussarbeit
- 124 • Vermittlung der Herangehensweise für die Strukturierung des wissen-schaftlichen  
125 Diskurses
- 126 • Entwicklung eines Grundverständnisses für die zu erstellende Abschlussarbeit

127 Nach der Freigabe und Besprechung des Exposé mit Ihrem Betreuer und dem  
128 Lehrstuhlinhaber, beginnen Sie damit, die eigentliche Arbeit zu schreiben und die

129 Planung, Konzeption und Implementierungen, die im Exposé bereits vorgeschlagen  
130 wurden, umzusetzen.

## 131 6. Die Abschlussarbeit

132 Durch die Erstellung des Exposé haben Sie bereits einen wichtigen Teil der Ab-  
133 schlussarbeit fertig gestellt: die saubere wissenschaftliche Verankerung, die daraus  
134 resultierende Ableitung der Forschungsziele, sowie den Ansatz und die Struktur der  
135 Arbeit. Deshalb können Sie die im Rahmen des Exposé erstellten Kapitel direkt in Ihre  
136 Abschlussarbeit übernehmen (bis einschl. 1.6). Nun erfolgt:

- 137 • eine weitergehende, selbständige Literaturrecherche
- 138 • eine Beschreibung des aktuellen Stands von Wissenschaft und Technik
- 139 • das Konzeptionelle Design der Architektur für die Umsetzung und Implemen-  
140 tierung
- 141 • die praktische Einarbeitung (Proof-Of-Concept Implementierungen, Kennenlernen  
142 der APIs und Frameworks)
- 143 • das Erstellen von Test-Szenarios, ggfs. auch die Organisation von Test-Daten
- 144 • die Implementierung
- 145 • eine Evaluierung des Ergebnisses mit Vergleich und Ausblick

146 Die jeweiligen Ergebnisse halten Sie in Ihrer Abschlussarbeit fest und folgen dabei  
147 analog zu den definierten Forschungszielen einer logischen Struktur:

148 **1. Einleitung** (folgt dem Exposé bis einschl. 1.6)

149 **2. Stand der Wissenschaft und Technik** (Für jedes Forschungsziel vom Typ Beobach-  
150 tung wird ein Kapitel vorgesehen und der zugehörige Stand der Wissenschaft  
151 beschrieben. Am Ende werden "remaining challenges" definiert und zusammenge-  
152 fasst.)

153 **3. Modellierung** (folgt dem Aufbau von Kapitel 2 und sieht für jedes Forschungsziel  
154 vom Typ Theoriebildung ein Kapitel vor)

155 **4. Implementierung** (folgt dem Aufbau von Kapitel 2 und bearbeitet Forschungsziele  
156 vom Typ Implementierung)

157 **5. Evaluierung** (folgt dem Aufbau von Kapitel 2 und bearbeitet Forschungsziele vom  
158 Typ Experiment)

159 **6. Diskussion** (Wie kann das Gesamtergebnis eingeordnet werden? Welche "remaining  
160 challenges" verbleiben? Was kann Teil von "future work" sein? Zusammenfassung)

161 Die Arbeit ist so ausgelegt, dass der Anteil von Theorie und Praxis jeweils 50%  
162 beträgt. Durch die hier bereits vorgeschlagene Herangehensweise werden Sie zu einer ef-  
163 fizienten, wissenschaftlich korrekten und nachvollziehbaren Arbeitsweise angeleitet. Im  
164 Rahmen der Betreuung am Lehrgebiet steht Ihnen ein Ansprechpartner zur Verfügung,  
165 der Ihnen sowohl technisch / fachlich, als auch strukturell Hilfestellung bieten kann. Mit  
166 dem Inhaber des Lehrgebiets werden die wichtigsten Meilensteine Ihrer Arbeit ebenfalls  
167 besprochen. Sollten Sie Interesse an diesem Thema haben, kontaktieren Sie bitte einen  
168 der genannten Ansprechpartner. Vielen Dank!

## References

1. Mc Kevitt et al. (2019), *Digital empathy secures Frankenstein's monster*; In CEUR Proc. of the 5th Collaborative European Research Conference (CERC 2019), Hochschule Darmstadt, University of Applied Sciences, Faculty of Computer Science, Darmstadt, Germany, March 29-30th, Vol-2348, 335-349
2. Beyerer J. (2017), *Pattern Recognition - Introduction*; Publisher: Walter de Gruyter GmbH & Co KG, Berlin; pp. 10-37, ISBN: 978-3-110-53794-9
3. Beierle C. (2019), *Methoden wissensbasierter Systeme - Grundlagen*; Publisher: Springer-Verlag, Berlin Heidelberg New York; pp. 89-157, ISBN: 978-3-658-27084-1

4. Goodfellow I. (2016), *Deep Learning*; Publisher: MIT Press, Cambridge; ISBN: 978-0-262-03561-3
5. Heaton J. (2015), *Deep Learning and Neural Networks*; pp. 1-25 ISBN: 978-1505714340
6. Google.com (2020), *Google Knowledge Search API*. Available online: <http://developers.google.com/knowledge-graph> (accessed 23.08.2020)
7. W3C.org (2020), *W3C Semantic Web Activity*. Available online: <http://w3.org/2001/sw> (accessed 23.08.2020)
8. Avola D. (2018), *Low-Level Feature Detectors and Descriptors for Smart Image and Video Analysis: A Comparative Study*; Journal: Intelligent Systems Reference Library; , pp 7-29; ISBN: 978-3-319-73890-1
9. Domingue J. (2011), *Introduction to the Semantic Web Technologies*. Available online: <https://doi.org/10.1007/978-3-540-92913-0> (accessed 2020)
10. Kwasnicka (2018), *Bridging the Semantic Gap in Image and Video Analysis*; Publisher: Springer, Berlin; pp. 97-118, ISBN: 978-3-319-73891-8
11. Mc Kevitt P. (2003), *MultiModal semantic representation*; Journal: In Proc. of the SIGSEM Working Group on the Representation of MultiModal Semantic Information, Editor Harry Bunt Kiyong Lee Laurent Romary and Emiel Kraemer
12. Spyrou (2017), *Semantic Multimedia Analysis and Processing*; Publisher: CRC Press, Boca Raton; pp.31-63, ISBN: 978-1-351-83183-3
13. Scherer R., *Computer Vision Methods for Fast Image Classification and Retrieval*; Publisher: Polish Academy of Science, Warsaw; pp. 33-134, ISBN: 978-3-030-12194-5
14. Mark Nixon, *Feature Extraction and Image Processing for Computer Vision*; Publisher: Academic Press Elsevir
15. Bhute B. (2012), *Multimedia Indexing and Retrieval Techniques: A Review*; Journal: International Journal of Computer Applications, Volume 58, pp 35-42
16. Smeulders A. (2000), *Content-based image retrieval at the end of the early years*; Journal: IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, pp 1349-1380
17. Gurski F. (2019), *On Characterizations for Subclasses of Directed Co-Graphs*. Available online: <http://arxiv.org/abs/1907.00801> (accessed 24.08.2020)
18. Needham M., *Graph Algorithms, Practical Examples in Apache Spark and Neo4j*; Publisher: O'Reilly; ISBN: 978-1-492-05781-9
19. Robinson I., *Graph Databases*; Publisher: O'Reilly; ISBN: 978-1-491-93089-2
20. Jiezhong Q. (2017), *Network Embedding as Matrix Factorization: Unifying DeepWalk*. Available online: <http://arxiv.org/abs/1710.02971> (accessed 24.09.2020)
21. Bai et al. (2019), *SimGNN: A Neural Network Approach to Fast Graph Similarity Computation*. Available online: <https://doi.org/10.1145/3289600.3290967> (accessed 24.11.2020)
22. W3C.org (2013), *SPARQL Query Language for RDF*. Available online: <https://www.w3.org/TR/sparql11-overview/> (accessed 23.08.2020)
23. Nkgau T. (2017), *Graph similarity algorithm evaluation*; Journal: 2017 Computing Conference, pp 272-278
24. Sluzek A. (2013), *Local Detection and Identification of Visual Data*
25. J. S. Sevak and A. D. Kapadia and J. B. Chavda and A. Shah and M. Rahevar (2017), *Survey on semantic image segmentation techniques*; pp 306-313
26. Arndt et al. (2008), *COMM: A Core Ontology for Multimedia Annotation*; Journal of Combinatorial Theory - JCT; , pp 403-421
27. Ni J et al. (2017), *Research on Semantic Annotation Based Image Fusion Algorithm*; Journal: 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC), pp 945-948
28. Gayathri N. (2019), *An Efficient Video Indexing and Retrieval Algorithm using Ensemble Classifier*; Journal: 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICECCOT), pp 250-258;
29. Draper N. (1986), *User Centered System Design - New Perspectives on Human-computer Interaction*; Publisher: Taylor & Francis, Justus-Liebig-Universität; ISBN: 978-0-898-59872-8
30. Nunamaker J. (1991), *Systems Development in InformationSystems Research*; Journal of Management Information Systems 1991, pp. 89-106
31. Fowler M. (2004), *UML Distilled - A Brief Guide to the Standard Object Modeling Language*; Publisher: Addison-Wesley Professional, Boston; ISBN: 978-0-321-19368-1
32. Planche B. (2019), *Computer Vision with TensorFlow 2*; Publisher: Packt Publishing; pp. 77-99, ISBN: 978-1-78883-064-5
33. Schmitt I. (2005), *WS-QBE: A QBE-Like Query Language for Complex Multimedia Queries*; Journal: 11th International Multimedia Modelling Conference, pp 222-229
34. Jung H. (2020), *Automated conversion from natural language query to SPARQL query*. Available online: <https://doi.org/10.1007/s10844-019-00589-2> (accessed 24.10.2020)
35. Bornschlegel M. (2016), *IVIS4BigData: A Reference Model for Advanced Visual Interfaces Supporting Big Data Analysis in Virtual Research Environments*; Journal: BDA@AVI
36. Wagenpfeil S. (2020), *Github Repository of GMAF and MMFVG*. Available online: <https://github.com/stefanwagenpfeil/GMAF/> (accessed 25.09.2020)
37. Wagenpfeil S. (2021), *Towards AI-bases Semantic Multimedia Indexing and Retrieval for Social Media on Smartphones*; SMAP 2020 Conference paper

38. Wagenpfeil S. (2021), *AI-Based Semantic Multimedia Indexing and Retrieval for Social Media on Smartphones*; <https://www.mdpi.com/2078-2489/12/1/43>
39. Wagenpfeil S. (2021), *Fast and Effective Retrieval for Large Multimedia Collections*; <https://www.mdpi.com/2504-2289/5/3/33>
40. Wagenpfeil S. (2021), *Towards Automated Semantic Explainability of Multimedia Feature Graphs*; <https://www.mdpi.com/2078-2489/12/12/502>