

**TIB** LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK

Hendrik Bunke, Ralph Ewerth, Junaid Ahmed Ghauri, Ralph Kölle, Eric Müller-Budack,  
Margret Plank, Christa Womser-Hacker

# Exploitation of Research Results through Visual Patent Retrieval (ExpResViP)

Sachbericht zum Verwendungsnachweis im BMBF Verbundprojekt

**FKZ:** 031O2004A

Hannover, Januar 2024

# Verbundpartner

TIB - Leibnizinformationszentrum Technik- und Naturwissenschaften (Hannover)

IWIST - Institut für Informationswissenschaft und Sprachtechnologie (Universität Hildesheim; im Unterauftrag der TIB)

LG - Leibniz-Gemeinschaft, Geschäftsstelle (Berlin)

IAIS - Fraunhofer-Institut für intelligente Analyse- und Informationssysteme (Sankt Augustin)

# Inhaltsverzeichnis

<b>1. Kurzbericht</b>	<b>3</b>
1.1 Aufgabenstellung	3
1.2 Stand von Wissenschaft und Technik	3
1.3 Ablauf und Zusammenarbeit	4
1.4 Ergebnisse des Vorhabens	5
<b>2. Ausführlicher Bericht</b>	<b>6</b>
2.1 Darstellung der durchgeführten Arbeiten	6
AP 1: Vorverarbeitung / Datenbestand	7
AP 2: Information Seeking Behavior	8
AP 3: Gestalt- und Funktionsanforderungen	9
AP 4: Grundlegende Methoden und Algorithmen der Bildanalyse	11
AP 5: Weiterführende Algorithmen und Methoden der Bildanalyse	15
AP 7: GUI Entwicklung / Formative Evaluation	18
AP 11: Projektmanagement	20
2.2 Mittelverwendung	21
2.3 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten	21
2.4 Nutzen und Verwertbarkeit	22
2.5 Projektrelevante Fortschritte Dritter	23
2.6 Publikationen	23
<b>3. Erfolgskontrollbericht</b>	<b>24</b>
3.1 Wissenschaftlich-technisches Ergebnis	24
3.2 Fortschreibung des Verwertungsplans	24
3.3 Arbeiten, die zu keiner Lösung geführt haben	25
3.4 Einhaltung der Ausgaben- und Zeitplanung	25
<b>Referenzen</b>	<b>26</b>
<b>Anlagen</b>	

# 1. Kurzbericht

## 1.1 Aufgabenstellung

Zur Beurteilung der Verwertungschancen von Erfindungen ist meist eine Recherche in der Patentliteratur nach vergleichbaren Technologien oder Verfahren erforderlich. Gerade für Forschungseinrichtungen können die Hürden für eine solche Recherche hoch sein. Sie verfügen in der Regel nicht über entsprechend geschultes Personal und für die Forscher:innen selbst ist der für eine erfolgreiche Stichwortsuche erforderliche Aufwand, sich in die hochspezifische und juristisch aufgeladene Patentsprache („Patentese“) einzuarbeiten, in der Regel kaum leistbar. Zu diesen fachlichen Sprachbarrieren kommen fremdsprachliche hinzu, wenn Verwertungschancen in verschiedenen Jurisdiktionen identifiziert werden sollen.

Ziel des Verbundprojektes ExpResViP war es daher, den Prototyp eines visuellen Patentinformationssystems (PIS) zu entwickeln. Mit diesem sollte es möglich sein, anhand von Bildern von Erfindungen in einer entsprechend aufbereiteten Datenbank nach semantisch ähnlichen Figuren in der Patentliteratur zu suchen, um Verwertungschancen leichter und effizienter zu erkennen. Mit dieser innovativen Funktion könnten Forschungsergebnisse schneller in Produkte und Märkte überführt und der Technologietransfer aus den Forschungseinrichtungen gestärkt werden.

Verbundpartner waren neben der TIB die Geschäftsstelle der Leibnizgemeinschaft (LG) und das Fraunhofer-Institut für intelligente Analyse- und Informationssysteme (IAIS). Als Unterauftragnehmer der TIB war beteiligt das Institut für Informationswissenschaft und Sprachtechnologie (IWIST) der Universität Hildesheim.

Wesentliche Aufgaben der TIB in diesem Verbund waren die Sammlung und Vorverarbeitung von Patentdaten aus den Bereichen *Autonomes Fahren* und *Wind-/Wasserturbinen*, die Entwicklung von grundlegenden und weiterführenden Methoden und Algorithmen der Bildanalyse, sowie das Projektmanagement. Der Unterauftrag an das IWIST umfasste die Analyse des Nutzer:innenverhaltens für den spezifischen Kontext, die Erstellung der daraus resultierenden Gestalt- und Funktionsanforderungen sowie die GUI Entwicklung des Prototyps.

## 1.2 Stand von Wissenschaft und Technik

Patentinformationssysteme basieren derzeit überwiegend auf der Einschränkung des Suchraums durch geeignete Filter und auf der Suche mit Stichwörtern. Das gilt besonders für die öffentlich zugänglichen PIS der Patentämter wie das DEPATISnet des Deutschen Patent- und Markenamts oder das Espacenet des Europäischen Patentamts. Einige nicht-öffentliche Anbieter von PIS bieten daneben auch die so genannte semantische Suche nach Patentschriften an. Bei

dieser wird in Patentdatenbanken nach Texten gesucht, die zu einer in natürlicher Sprache formulierten Suchanfrage semantisch möglichst ähnlich sind.

Anwendungen, mit denen die in den Patentabbildungen enthaltenen Informationen extrahiert, verarbeitet und der Patentrecherche zugänglich gemacht werden, gibt es dagegen bisher erst in Ansätzen. So verknüpft zum Beispiel Google Patents die Referenzzeichen in den Patentabbildungen (teilweise) mit den entsprechenden textlichen Beschreibungen. Ein öffentlich zugängliches PIS, das den vollen Informationsgehalt von Patentabbildungen für die Patentrecherche erschließt, gibt es dagegen bisher nicht. Besonders die Suche nach relevanten Patentschriften auf Basis von Bildern, die Nutzer:innen hochladen können, bietet nach unserer Kenntnis bisher kein PIS an.

Ein wichtiger Grund für diese Technologielücke ist der besondere Charakter von Patentabbildungen. Nach den Richtlinien der Patentämter sind sie als farblose Strichzeichnungen anzufertigen und mit Referenzzeichen (Nummern) zu versehen, die auf bestimmte Details innerhalb einer Figur verweisen. Erst jüngste Fortschritte in den Bereichen Deep Learning und Computer Vision erlauben es, diese so zu verarbeiten, dass sowohl so genannte low-level Merkmale wie Formen als auch höhere visuelle Konzepte erkannt und für die hier wichtigen semantischen Analysen genutzt werden können. Vor allem neuere Ansätze, die visuelle und textuelle Informationen gleichzeitig verarbeiten können, wie beispielsweise das Contrastive Language-Image Pre-training, sind hier von besonderem Interesse.

### 1.3 Ablauf und Zusammenarbeit

Durch die Corona-Pandemie kam es zunächst zu Problemen bei der Stellenbesetzung und der Organisation der Zusammenarbeit. Auch im weiteren Verlauf mussten alle Projektpartner mit Personalfluktuationen umgehen. Dennoch wurde in den einzelnen Arbeitspaketen kontinuierlich gearbeitet. Sehr früh waren bereits die Patentbestände aus den Domänen *Autonomes Fahren* und *Windkraftanlagen* gesichtet und aufbereitet, die dann als Datengrundlage des Projekts dienten. Durch Workshops, regelmäßige Videokonferenzen und ein gemeinsames Wiki wurde die Kooperation organisiert. Inhaltlich ergaben sich im Vergleich zur Planung ebenfalls Änderungen, insbesondere durch neue Voraussetzungen im momentan sehr dynamischen Themenkomplex *Künstliche Intelligenz*. So ergab sich durch die Einführung des Transformer-Modells CLIP (Contrastive Language-Image Pretraining; Radford et al., 2021) eine neue, sehr vielversprechende Forschungsrichtung, die wir in unsere Arbeit integrierten. Im letzten Projektdrittel wurden die Ergebnisse dann in Form des Prototyps *VisPat* (Visual Search for Patents) zusammengeführt. Für die Entwicklungskooperation wurden weitere Strukturen (Gitlab, Jira, wöchentliche Developermeetings) geschaffen, so dass es trotz der zeitlichen Verzögerungen und personellen Probleme gelang, einen lauffähigen Prototypen zu erstellen, der die wesentlichen Ziele des Projekts demonstrieren kann. Der im Projekt erarbeitete Code wird so weit wie möglich veröffentlicht.

## 1.4 Ergebnisse des Vorhabens

Mit dem Tool *VisPat* (Visual Search for Patents) konnte das Projekt den funktionsfähigen Prototyp eines visuellen Patentinformationssystems vorlegen.<sup>1</sup> Dieser bietet die angestrebte innovative Bildsuche zusammen mit einer semantischen textuellen Suchfunktion und der herkömmlichen Stichwortsuche auf einer einfachen Benutzer:innenoberfläche an. *VisPat* ist grundsätzlich in der Lage, auf Basis eines Suchbildes in der angeschlossenen Patentdatenbank ähnliche visuelle Konzepte in Patentabbildungen zu erkennen und so Patentedokumente auszugeben, die für Nutzer:innen relevant sind. Diese Funktionalität stellt ein Alleinstellungsmerkmal gegenüber existierenden Patentsuchsystemen dar, die sich vorrangig auf die Suche mithilfe von Metadaten oder Textinformationen beschränken.

Anders als geplant ist ein Produktiveinsatz von *VisPat* allerdings derzeit nicht oder nur beschränkt möglich. Ein Grund ist dafür die geringe Abdeckung der Patentliteratur in der angeschlossenen Datenbank (ca. 34.000 Patente und 350.000 Abbildungen in Patentklassen, die die exemplarischen Technologiebereiche „Autonomes Fahren“ und „Windkraftanlagen“ beinhalten). Auch konnte nicht abschließend geklärt werden, für welche Rechercheanlässe die eingesetzten Technologien zuverlässige Ergebnisse liefern können. Wie eine qualitative Evaluation gezeigt hat, deutet sich an, dass *VisPat* etwa für Neuheits- und Patentierbarkeitsrecherchen nur begrenzt nutzbar ist. Übersichtsrecherchen zum Stand der Technik kann das Tool dagegen sinnvoll unterstützen. Es zeigte sich außerdem, dass die eingesetzten Deep-Learning-Modelle wesentliche visuelle Konzepte in Patentabbildungen noch nicht optimal repräsentieren können. Trotz dieser Beschränkungen und einem zwar funktionalen, aber noch nicht visuellen Standards entsprechenden GUI waren erste Rückmeldungen von Patentexperten positiv.

Insgesamt hat das Projekt gezeigt, dass eine Erweiterung von Patentinformationssystemen um visuelle Such- und Analysefunktionen die Patentrecherche wirksam unterstützen kann. Entsprechende Anwendungen wie *VisPat* müssen jedoch weiter verbessert werden, um in der Transferpraxis von Forschungseinrichtungen sinnvoll eingesetzt werden zu können. Dafür konnte ExpResViP mehrere, viel versprechende Optimierungsoptionen identifizieren, auf denen anwendungsorientierte Folgevorhaben und die einschlägige Forschung aufsetzen können.

---

<sup>1</sup> <https://service.tib.eu/vispat>

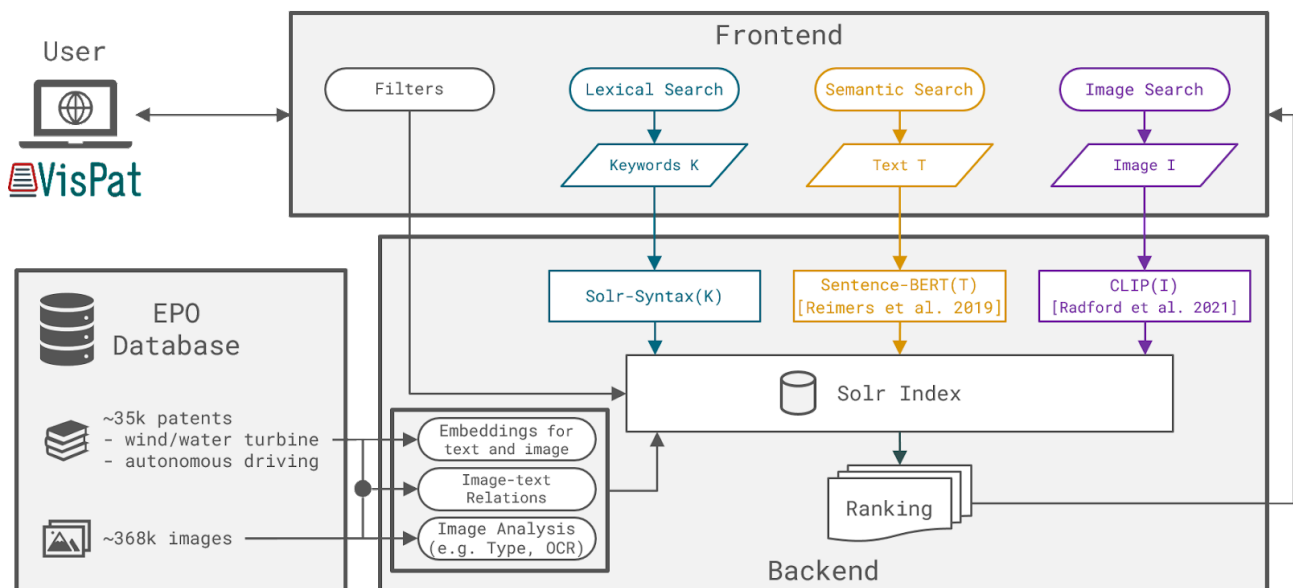
## 2. Ausführlicher Bericht

### 2.1 Darstellung der durchgeführten Arbeiten

Die durchgeführten Arbeiten werden hier beschrieben anhand der Arbeitspakete des Projekts, für die die TIB verantwortlich war oder an denen sie beteiligt war (vgl. Tabelle 1). Die Arbeitspakete 2 und 7 wurden im Auftrag der TIB vom *Institut für Informationswissenschaft & Sprachtechnologie* (IWIST) der Universität Hildesheim umgesetzt. Die *Summative Evaluation* (AP 9) wurde, anders als vorgesehen, durch den Projektpartner *Leibniz-Gemeinschaft* durchgeführt. Dieser unterstützte in der Endphase des Projekts auch die Arbeiten im Projektmanagement (AP 11). Die Gründe für diese Änderungen ergeben sich aus den nachfolgenden Tätigkeits- und Prozessbeschreibungen der einzelnen Arbeitspakete.

AP	Titel	Verantwortlich	Beteiligt
1	Vorverarbeitung / Datenbestand	TIB	IAIS, LG
2	Information Seeking Behavior	TIB	IWIST, IAIS, LG
3	Gestalt- und Funktionsanforderungen	TIB	IWIST, IAIS
4	Grundlegende Methoden und Algorithmen der Bildanalyse	TIB	
5	Weiterführende Methoden und Algorithmen der Bildanalyse	TIB	IAIS
6	Methoden der semantischen Suche	IAIS	
7	GUI Entwicklung / Formative Evaluation	TIB	IWIST, IAIS, LG
8	Systemintegration zum Patentretrieval Tool	IAIS	TIB, IWIST, LG
9	Summative Evaluation	LG	
10	Kompetenzaufbau und Transfer	LG	TIB, IAIS
11	Projektmanagement	TIB, LG	IAIS

**Tab. 1:** Arbeitspakete und Verantwortlichkeiten



**Abb. 1:** Architektur des entwickelten Prototypen *VisPat* zur textuellen und visuellen Patensuche.

## AP 1: Vorverarbeitung / Datenbestand

Ziel dieses Arbeitspakets (AP) war die Sammlung und Vorverarbeitung von Patentdaten aus den Bereichen autonomes Fahren und Wind-/Wasserturbinen. Zunächst wurden für beide Domänen relevante Klassen nach dem IPC (International Patent Classification) Klassifikationsschema identifiziert. Auf Basis dieser Klassen wurden 41.730 Patente aus beiden Domänen (circa 20.000 Patent pro Domäne) über die öffentlich zugängliche Schnittstelle des Publikationsserver<sup>2</sup> des Europäischen Patentamtes gesammelt. Bei der Vorverarbeitung der Rohdaten wurden die dazugehörigen PDF und XML-Dateien sowie die Bilddaten der Patente extrahiert und circa 7.000 Patente mit unvollständigen Informationen (d.h. dass eine dieser drei Eingaben fehlt) entfernt. Aus den XML-Dateien wurden Metainformationen (z.B. Titel, Anmelder, IPC Klassifikation, Sprache, Datum der Veröffentlichung) und textuellen Informationen (z.B. Zusammenfassung, Patentansprüche, Volltext) der Patente extrahiert und zusammen mit den Bilddaten in einer Datenbank (Apache Solr<sup>3</sup>) gespeichert. Der resultierende Datenbestand nach der Vorverarbeitung umfasst 34.754 Patente und insgesamt 368.971 Bilder. Eine detaillierte Statistik über den gesammelten Datenbestand ist in Tab. 2 dargestellt. Die Umsetzung des AP erfolgt wie geplant.

IPC	Domäne	#Patente	#Bilder
<a href="#">B60W</a>	CONJOINT CONTROL OF VEHICLE SUB-UNITS OF DIFFERENT TYPE OR DIFFERENT FUNCTION; CONTROL SYSTEMS SPECIALLY ADAPTED FOR HYBRID VEHICLES; ROAD VEHICLE DRIVE CONTROL SYSTEMS FOR PURPOSES NOT RELATED TO THE CONTROL OF A PARTICULAR SUB-UNIT	6.892	82.315
<a href="#">F01D 5</a>	Blades; Blade-carrying members (nozzle boxes F01D 9/02);	3.480	18.837

<sup>2</sup> <https://data.epo.org/publication-server/>

<sup>3</sup> <https://solr.apache.org/>

	Heating, heat-insulating, cooling, or antivibration means on the blades or the members		
<a href="#">F03D</a>	WIND MOTORS	4.679	37.889
<a href="#">G05D</a>	SYSTEMS FOR CONTROLLING OR REGULATING NON-ELECTRIC VARIABLES	4.794	53.893
<a href="#">G06F</a>	ELECTRIC DIGITAL DATA PROCESSING	5.905	76.842
<a href="#">H02J</a>	CIRCUIT ARRANGEMENTS OR SYSTEMS FOR SUPPLYING OR DISTRIBUTING ELECTRIC POWER; SYSTEMS FOR STORING ELECTRIC ENERGY	9.004	99.195
<b>Total</b>		<b>34.754</b>	<b>368.971</b>

**Tab. 2:** Anzahl an gesammelten Patenten und Bildern für relevante Klassen der Domänen Wind- und Wasserturbinen sowie autonomes Fahren nach dem IPC (International Patent Classification) Klassifikationsschema

## AP 2: Information Seeking Behavior

In diesem Arbeitspaket ging es darum, eine Wissensgrundlage zu schaffen, um die angestrebte Benutzer:innenorientierung des späteren Prototyps und dessen Benutzer:innenschnittstelle (GUI) sicherzustellen. Dazu wurden zunächst die Charakteristika der Patentinformation, die bestehenden Patentinformationssysteme etc. zusammen mit der Leibnizgemeinschaft genau analysiert. Arbeitsumfeld und Informationsverhalten der Benutzenden – insbesondere in Bezug auf die visuelle Suche – wurden erhoben. Dazu musste ein Kreis an im Patentbereich arbeitenden Experten und Expertinnen rekrutiert werden. Insgesamt ist festzustellen, dass es sehr viele verschiedene Aufgabenbereiche in dieser Sparte gibt, die hier Berücksichtigung finden sollten.

Die Literatur aus den verschiedenen wissenschaftlichen Bereichen wurde als Ausgangspunkt herangezogen. Eine gut geeignete Übersicht über Aufgaben und Szenarios im Patentretrieval findet sich in Shalaby & Zadrozny 2019 sowie in Azzopardi et al. 2010:

Search task	Who	When	Purpose	Output
Related work	Inventor/prosecutor	Pre-grant	All related work	Applicant's disclosure
Patentability	Prosecutor/examiner	Pre-grant/examination	Novelty breaking work	Grant/modify/reject
Infringement	Owner/investor	Post-grant	Relevant claims/infringing products	Sue/license/clearance
Freedom to operate	Investor	Post-grant	Relevant claims/related work	Clearance
Invalidity	Competitor/defendant	Post-grant	Novelty breaking work	Reexamine/inter-parts review/post-grant review
Technology survey	Technology analyst	Pre/post-grant	All published patents	Survey report

**Tab. 3:** Scenarios of Prior Art Search, p. 638

Methodisch kamen leitfragenorientierte Experteninterviews sowie Fragebögen zum Einsatz, um das Informationsverhalten sowie den jeweiligen Kontext der Personen erschließen zu können. Hier zeigte sich, dass aufgrund der Sensibilität des Patentbereichs spezielle Datenschutzrichtlinien einzuhalten waren, Interviewverträge geschlossen und z.T.

Verschwiegenheitserklärungen abgegeben werden mussten. Dennoch erwiesen sich die beteiligten Expert:innen als sehr offen und waren gerne bereit, dem Projekt ihre Expertise und ihr Wissen zur Verfügung zu stellen.

Zunächst wurden Leitfragen generiert, die mit den Projektpartnern des Retrieval- und Datenbankteams sowie den Bilderkennern diskutiert wurden. Zentral war in diesem Zusammenhang die Rolle von Abbildungen sowie die Verknüpfungen zwischen Text und Bild bei der Suche nach Patenten/Offenlegungen oder bei der Prüfung einer entsprechenden Ergebnismenge. Dies mündete in eine Agenda, die sowohl in den Interviews als auch bei den Fragebögen eingesetzt wurde (s. Anlage 1). Insgesamt konnten elf Interviewpartner:innen gewonnen werden.

- Prüfer aus den Patentämtern: 3 Personen
- Patentanwältin: 1
- Patentspezialist:in aus der Industrie: 2
- Erfinder: 1
- Transfer- und Beratungsinstitutionen: 4

Weitere fünf Personen beantworteten den Fragebogen.

Nachdem die Fragestellungen in einem Pretest einen Test durchlaufen hatten, wurden die Interviews über eine geschlossene Online-Plattform durchgeführt und dauerten jeweils ein bis zwei Stunden. Von Projektseite waren drei Personen beteiligt, die verschiedene Aufgaben wahrnahmen (Führung und Moderation, inhaltliche Fragen und Notizen, Mitschrift). Die Interviews wurden (manuell und automatisch) transkribiert und anschließend einer qualitativen Inhaltsanalyse (s. Mayring 2015, Kuckartz 2018) unterzogen. Dafür wurde ein Kategoriensystem (induktiv/deduktiv) (s. Anlage 2) entwickelt, dem die Aussagen der Interviewpartner:innen zugeordnet wurden. Der Rücklauf der Fragebögen wurde gleichermaßen ausgewertet, erwies sich aber als weit weniger ergiebig.

Aus dieser Prozedur konnten zwei sog. Super-User heraus destilliert werden, die sich bereit erklärten, das Projekt und die Prototypentwicklung auch weiterhin zu unterstützen und zu beraten (z.B. im Bereich der formativen Evaluation und der Testaufgabenentwicklung).

### AP 3: Gestalt- und Funktionsanforderungen

Die Analyseergebnisse des Expertenverhaltens bilden die Basis für die Identifikation des Funktionsumfangs und die Darstellungsform bei der Prototypentwicklung. Sie dienen auch dazu, bei der Evaluierung auf authentisches Material zurückgreifen zu können.

Aus dem in Kategorien gefassten Wissen konnte auf diese Weise ein Set an Anforderungen entwickelt werden, die sich an verschiedene Teams des Projektes richteten. Die Anforderungen

wurden in natürlicher Sprache in einem bestimmten Format formuliert und mit allen Projektpartnern diskutiert und priorisiert. Die Anforderungen wurden von den Teams einer Analyse unterzogen, die die Machbarkeit einschätzte. Nicht alle Anforderungen waren spezifisch genug, um eine Realisierung in vollem Umfang während der Projektlaufzeit zu ermöglichen.

Beispiel:

ID	Anforderung
R1	<p>Das System muss den Nutzer:innen die Möglichkeit bieten, eine Suchanfrage sowohl nur mit Bildern als auch in Kombination mit Stichwörtern und/oder Klassen durchzuführen.</p> <p><b>Beschreibung:</b> Es muss die Möglichkeit geben, mit Bildern zu suchen. Entsprechend muss das System das Hochladen eines Bildes zur Verfügung stellen (bspw. wie bei Google-Bildersuche).</p> <p>Es muss weiterhin die Möglichkeit geben, die Bildersuche mit einer Textsuche zu kombinieren. Es muss den Nutzer:innen möglich sein, dem Bild Stichwörter, Klassen, Phrasen etc. beizufügen.</p>

Eine vollständige Liste der Anforderungen findet sich in Anlage 3.

Im Anschluss daran wurden verschiedene Lösungsansätze für die Benutzer:innenschnittstelle (GUI) des Systems entwickelt und mit Probanden getestet. Die Vorgehensweise folgte dem User Centered Design Ansatz und bezog die Benutzenden in jeder Phase ein:

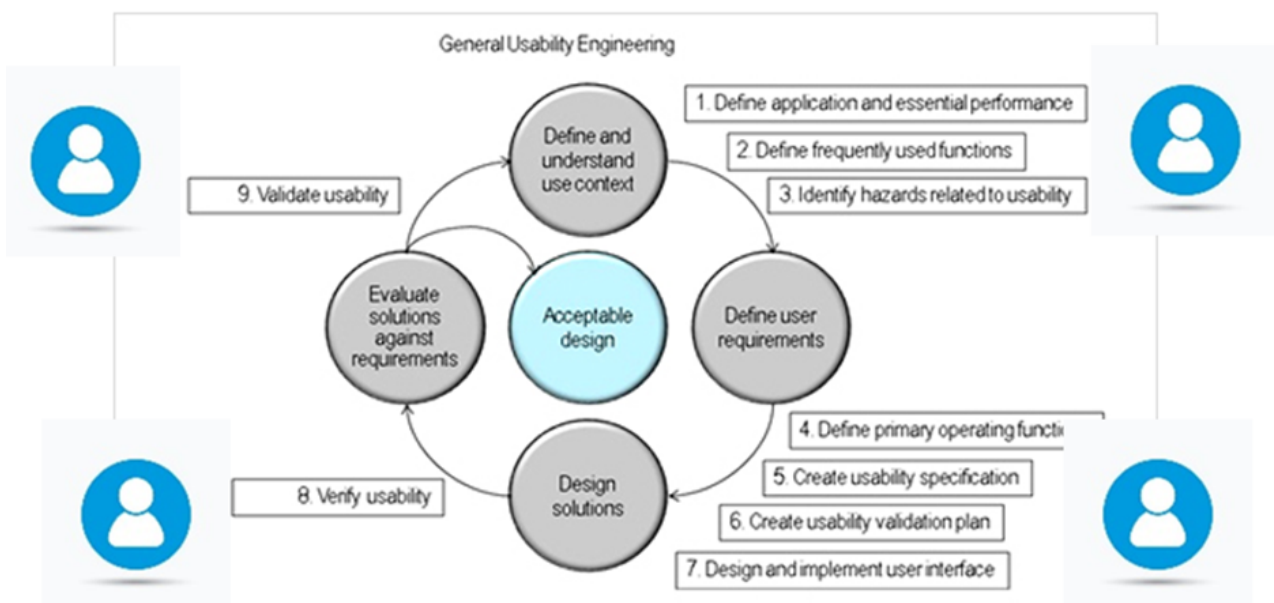


Abb. 2: User Centered Design Prozess

Zwei der so entstandenen GUI-Prototypen wurden in kurzen, kompakten Videos mit Erklärungen und Fragestellungen den Expert:innen präsentiert. Für diese stellte es eine große Erleichterung

dar, dass sie die Prototypen zeitunabhängig betrachten und bewerten konnten. Alle wählten die Option einer textuellen Antwort, obwohl auch eine gesprochene Einschätzung während der Betrachtung möglich gewesen wäre. Die gegebenen Antworten waren zwischen einer und sieben DIN A4-Seiten lang und stellten eine wichtige Quelle für die Modifikation der Prototypen dar. Die Prototypen mussten z.B. terminologische, inhaltliche, ästhetische Aspekte in die nächste Iteration einbeziehen.

Die zweite Iteration des GUI-Prototypen musste etwas zurückgestellt werden, da die Systemintegration einen großen Stellenwert einnahm. Die lexikalische und die semantische Textsuche sowie die Koordination mit den Abbildungen musste zunächst erarbeitet werden, damit eine Operationalisierung auf der GUI möglich war.

Die Videos und die Bewertungen der Expert:innen können auf Anfrage als Materialien zur Verfügung gestellt werden.

## AP 4: Grundlegende Methoden und Algorithmen der Bildanalyse

In diesem Arbeitspaket sollten grundlegende Methoden beispielsweise zur Erkennung typischer Abbildungsformen in Patenten, von Texten und Zahlen in Abbildungen, sowie zur automatischen Erkennung von Bild-Text-Bezügen entwickelt werden.

Für die **Klassifizierung typischer Abbildungsformen und Bildperspektiven** wurden zunächst passende Datensätze in der Literatur zum Training von Deep-Learning-Modellen gesucht. Der CLEF-IP Datensatz (Piroi et al., 2011) enthält neun relevante Klassen (siehe Tab. 2) für das Projektvorhaben. Auf Basis der Projektanforderungen wurde der Datensatz mit einer Klasse für Block- und Schaltdiagramme erweitert. Die Bilder dieser Klasse haben wir zunächst mithilfe eines regelbasierten Ansatzes aus unserem Datenkorpus (AP 1) extrahiert und anschließend zur Sicherstellung der Güte mittels eines eigens entwickelten Annotationstools manuell annotiert. Wei et al. 2022 haben mithilfe der Bildunterschriften von Patentbildern aus dem *United States Patent and Trademark Office* (USPTO) automatisch Metainformationen extrahiert, die unter anderem die Bildperspektive enthalten. Auf Basis dieses Datensatzes haben wir Klassen für Bildperspektiven in drei Hierarchiestufen identifiziert und die Bilddaten automatisch über die extrahierten Metadaten gelabelt. Die Statistiken beider Datensätze können Tab. 4 und Tab. 5 entnommen werden. Verschiedene Deep-Learning-Modelle auf Basis von künstlichen neuronalen Faltungsnetzen (engl.: Convolutional Neural Networks, CNNs) sowie eines aktuellen Transformer-Modells (Radford et al., 2021) wurden mit geeigneten Klassifikationsschichten erweitert und auf Basis dieser Datensätze trainiert. Die besten Ergebnisse wurden mit der ResNeXt-Architektur (Xie et al., 2017) erzielt. Konfusionsmatrizen für beide Aufgabenstellungen sind in Abb. 3 dargestellt. Die Ergebnisse mit 85% Makrogenauigkeit bei der Klassifikation von Abbildungsformen und 93% bei der Klassifikation von Bildperspektiven (Hierarchieebene 1) sind insgesamt sehr positiv zu bewerten. Die Ausgaben dienen insbesondere zur Facettierung der Suche in der im Projekt entwickelten prototypischen Applikation (*VisPat*). Die Arbeit wurde bei

der International Conference on Theory and Practice of Digital Libraries (TPDL) 2023 vorgestellt (Ghuri et al., 2023).

Abbildungstyp	Anzahl Bilder für		
	Training	Validierung	Test
Block- und Schaltdiagramme	450	50	100
Chemische Strukturen	5.362	595	112
Zeichnung	5.009	556	274
Flowchart	279	31	102
Gensequenzen	5.385	598	24
Graphen	1.497	166	193
Mathematische Formeln	5.355	595	126
Program	5.016	557	26
Symbol	1.421	157	17
Tabelle	4.952	550	66

**Tab. 4:** Statistiken für den CLEF-IP 2011 Datensatz (Piroi et al., 2011) erweitert mit einer Klasse für Block- und Schaltdiagramme

Bildperspektive	Anzahl Bilder für			Hierarchiestufe		
	Training	Validierung	Test	1 (2 Klassen)	2 (4 Klassen)	3 (7 Klassen)
● Perspektivansicht	6.140	150	150	yes	yes	yes
● Planare Ansicht	18.740	900	900	yes	no	no
● Links-Rechts	4.767	300	300	no	yes	no
● Links	2.407	150	150	no	no	yes
● Rechts	2.360	150	150	no	no	yes
● Unten-Oben	6.060	300	300	no	yes	no
● Unten	2.800	150	150	no	no	yes
● Oben	3.260	150	150	no	no	yes
● Vorne-Hinten	7.643	300	300	no	yes	no
● Vorne	5.184	150	150	no	no	yes
● Hinten	2.459	150	150	no	no	yes

**Tab. 5:** Statistiken für den USPTO-PIP Datensatz (United States Patent and Trademark Office - Patent Image Perspective) zur Klassifikation von Bildperspektiven mit drei Hierarchiestufen auf Basis von Wei et al. 2022



Die Erkennung von Referenzen (in der Regel in Form von Zahlen) in den Patentbildern, ermöglicht die **automatische Identifikation von bildbeschreibenden Textstellen** im Dokument. Hierfür werden zunächst sogenannte “Stop Words” und Sonderzeichen aus dem Text mithilfe der NLTK (Natural Language Toolkit) Python-Bibliothek<sup>7</sup> entfernt. Anschließend werden mithilfe von regulären Ausdrücken alle Zahlen und die dahinterliegenden Label (Name der Komponente) in den Textinformationen des Patentes extrahiert. Durch ein Mapping der detektierten Zahlen in Bild (durch OCR) und Text werden automatisch Bild-Text-Bezüge in Patenten bestimmt. Wie in Abb. 5 dargestellt, können die Bild-Text Bezüge die Nutzer:innen bei der Exploration von Patenten unterstützen, da über das Frontend direkt Referenzen im Bild aufgelöst werden und zudem die relevanten Textpassagen zu einem Bild gefiltert werden können.

The screenshot shows the VisPat interface for patent EP09734051B1. The top navigation bar includes 'BIBLIOGRAPHIC DATA', 'DESCRIPTION', 'CLAIMS', 'DRAWINGS', and 'CITATIONS'. The 'DESCRIPTION' section is active, showing a search for 'wind turbine 102'. A dropdown menu lists search results: '106 nacelle', '110 blades', '108 hub', '104 tower', and '102 wind turbine'. The main text area contains a paragraph describing the wind turbine 102, with terms like 'tower 104', 'nacelle 106', 'hub 108', 'blades 110', and 'wind turbine 102' highlighted. To the right, a technical drawing of the wind turbine 102 is shown with numbered labels (102, 104, 106, 108, 110, 114, 116) corresponding to the components mentioned in the text.

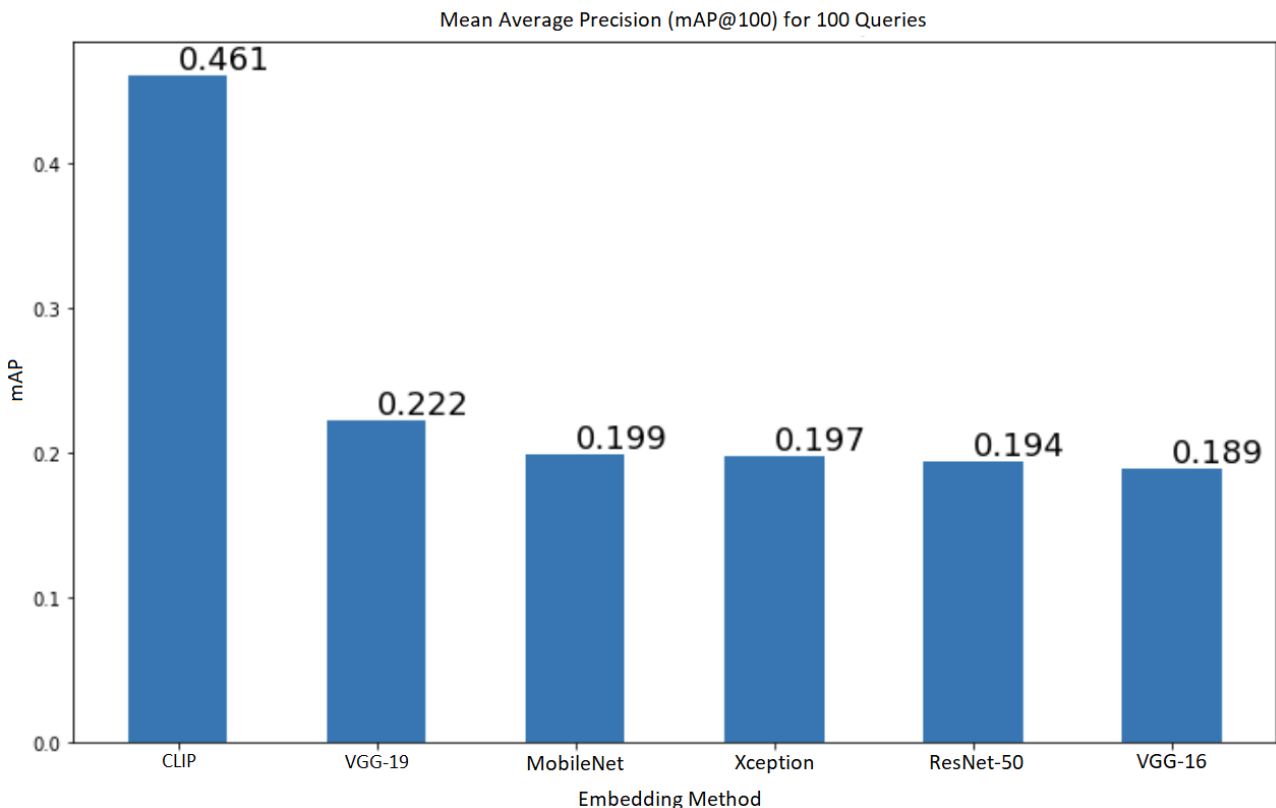
**Abb. 5:** Darstellung von Bild-Text-Bezügen im Frontend von VisPat. Links: Textreferenzen des Konzeptes “wind turbine 102”. Rechts: Bild und extrahierte Komponenten aus dem Text.

Die Detektion von Pfeilen und Linien wurde auf Basis der Arbeit von Kembhavi et al. 2016 implementiert und qualitativ ausgewertet. Ziel war die Herstellung von Bild-Text-Bezügen, die allerdings in höherer Güte bereits durch das zuvor vorgestellte Verfahren des Mappings von Referenzen zwischen Text und OCR-Ausgaben hergestellt werden konnten. Deshalb haben wir uns auf die Verbesserung dieses Ansatzes fokussiert und die Detektion von Pfeilen und Linien nicht weiter verfolgt.

<sup>7</sup> <https://www.nltk.org/>

## AP 5: Weiterführende Algorithmen und Methoden der Bildanalyse

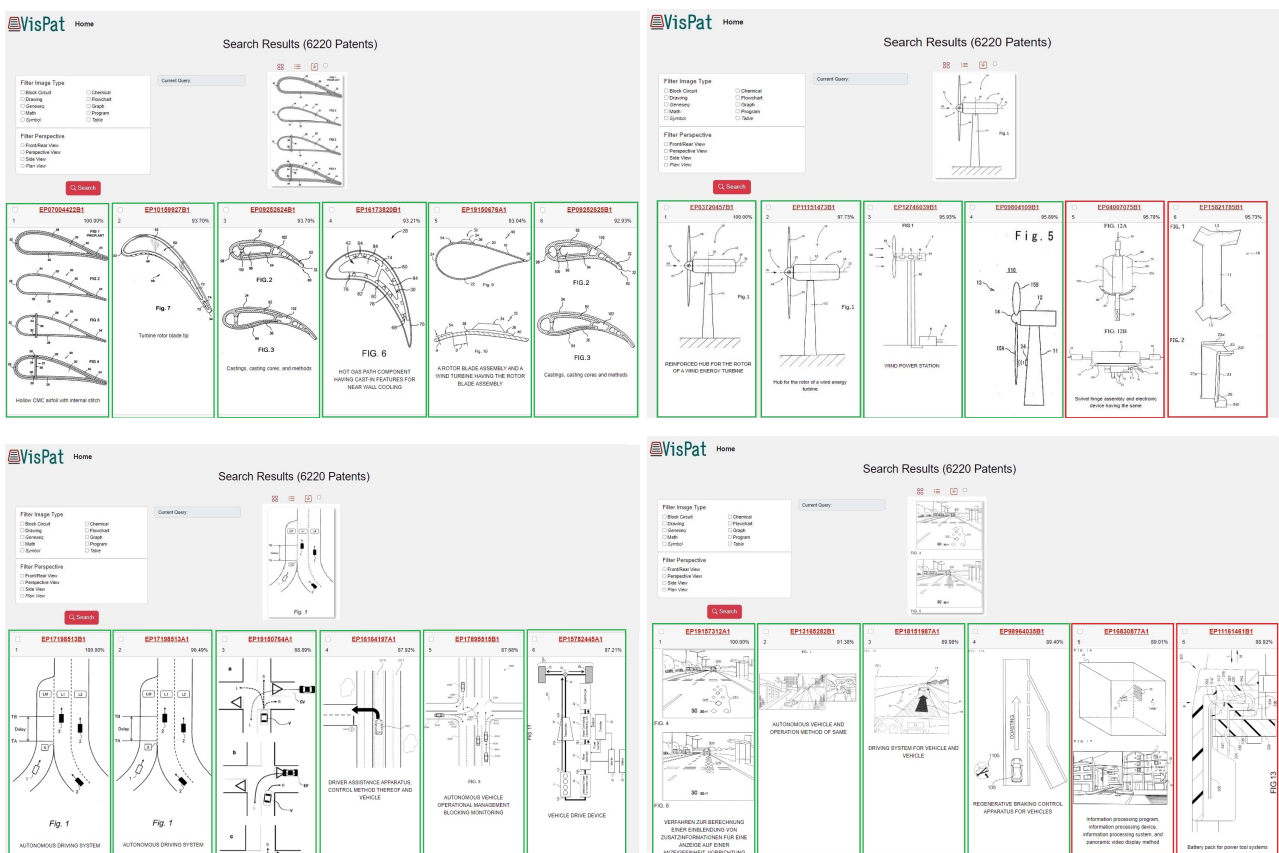
Das primäre Ziel dieses Arbeitspakets war die Entwicklung eines Plattformmoduls zur bildbasierten Ähnlichkeitssuche. Hierfür sollten Methoden zur Erkennung von visuellen Objekten und Konzepten auf Basis von Lexika für verschiedene thematische Schwerpunkte (hier: Wind- und Wasserturbinen sowie autonomes Fahren) entwickelt werden. Durch die Einführung des Transformer-Modells CLIP (Contrastive Language-Image Pretraining; Radford et al., 2021) während der Bearbeitung dieses Vorhabens hat sich jedoch eine neue, sehr vielversprechende Forschungsrichtung ergeben. Zunächst haben wir verschiedene Deep-Learning-Modelle (siehe Abb. 6) mit CLIP hinsichtlich der bildbasierten Ähnlichkeitssuche verglichen. Hierfür haben wir einen Testdatensatz mit 100 Query-Bildern erstellt und jeweils die Average Precision für die Top-100 Retrievalergebnisse berechnet. Die Ergebnisse für die mean Average Precision (mAP) über alle Queries haben gezeigt, dass CLIP bisherigen Modellen deutlich überlegen ist.



**Abb. 6:** Retrieval-Performance hinsichtlich der mean Average Precision (mAP) verschiedener Deep-Learning-Verfahren für die bildbasierte Ähnlichkeitssuche auf Basis von 100 Bildanfragen

Deshalb haben wir den Vision Transformer von CLIP implementiert, um semantische Informationen der Bilder in unserem Datenkorpus (AP 1) zu extrahieren und in einem Solr-Index abzuspeichern. Wie in Abb. 1 dargestellt, wurde ein Webdienst auf Basis von Flask mithilfe einer REST API (Representational State Transfer Application Programming Interface) zur **bildbasierten Ähnlichkeitssuche** entwickelt. Auf Basis eines hochgeladenen Eingabebildes werden zunächst semantische Bildmerkmale mit CLIP berechnet. Anschließend werden diese Merkmale auf Basis der Kosinusähnlichkeit mit den Bildmerkmalen des Korpus im Solr-Index (AP 1) verglichen, um

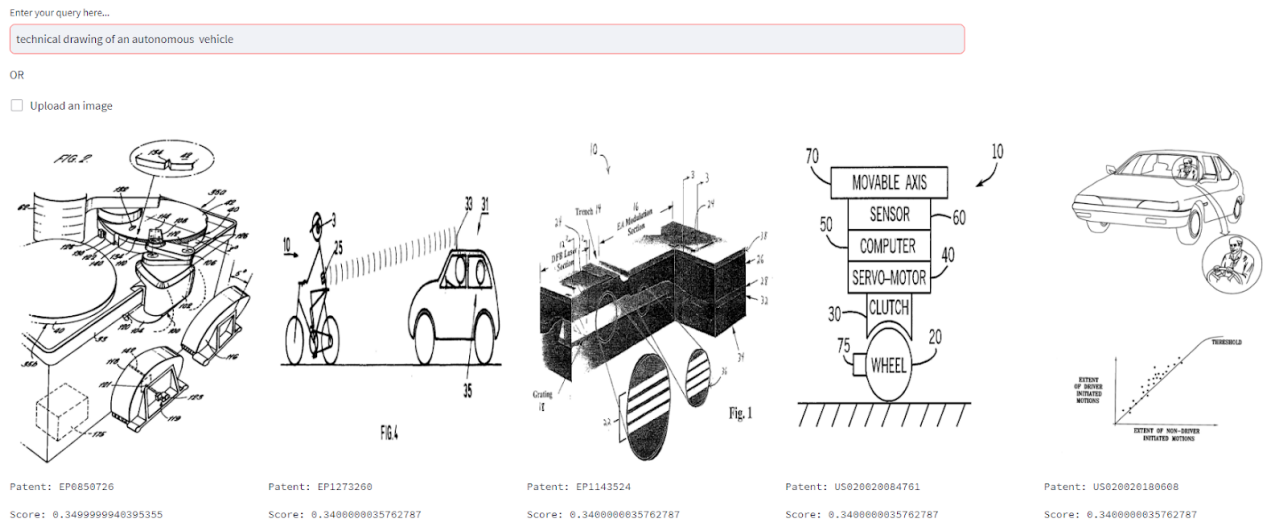
ein Ranking zu erstellen. Wir haben unseren Ansatz mithilfe des CLEF-IP 2011 Benchmarks (Piroi et al., 2011) zum bildbasierten Patentretrieval (Task: IMG-PAC) evaluiert und eine mean Average Precision (mAP) von 0,34 erzielt. Die bisher besten Ergebnisse lagen bei 0.27 (Piroi & Hanbury, 2017). Zudem wurde die bildbasierte Suche von den Projektpartnern qualitativ evaluiert und der entwickelte Ansatz hat in vielen Fällen vielversprechende Ergebnisse erzielt (siehe Abb. 7). Allerdings konnten so auch Schwachstellen identifiziert werden. Beispielsweise spielen Ähnlichkeiten hinsichtlich des prinzipiellen Layouts einer Abbildung (z.B. technische Zeichnung) oft eine größere Rolle als semantische Ähnlichkeiten, und die Erklärbarkeit, insbesondere bei irrelevanten Ergebnissen, ist durch die Anwendung eines Deep-Learning-Systems limitiert. Auf Basis dieser Erkenntnisse wurden mögliche Ansätze zur Verbesserung identifiziert. Beispielsweise könnten durch eine Feinanpassung (Optimierung) des CLIP-Modells an die Patentdomäne vermutlich wesentlich bessere Ergebnisse erzielt werden.



**Abb. 7:** Qualitative Ergebnisse der bildbasierten Ähnlichkeitssuche mit relevanten (grün) und irrelevanten Treffern (rot). Oben: Beispiele für ein Anfragebild eines Windradflügels und eines Windrads. Unter: Beispiele aus dem Bereich des autonomen Fahrens zu Spurhalteassistenten.

CLIP wurde mit circa 400 Millionen Bild-Text-Paaren in einem kontrastiven Trainingsprozess optimiert und extrahiert Bild- und Textmerkmale in einem gemeinsamen Merkmalsraum. Deshalb ermöglicht CLIP im Gegensatz zu CNN-basierten Ansätzen auch eine **intermodale Ähnlichkeitssuche**. Dabei können ähnliche Bilder auf Basis von Benutzer:innendefinierten Textbeschreibungen (als Prompts bezeichnet) gefunden werden. Die TIB hat hierfür geeignete Backend-Schnittstellen entwickelt, die perspektivisch in VisPat integriert werden könnten. Über

eine lokale Demo kann die Funktionalität bereits jetzt verwendet werden. In Abb. 8 sind qualitative Ergebnisse dargestellt.



**Abb. 8:** Qualitatives Ergebnis einer intermodalen Suche auf Basis des Prompts “technical drawing of an autonomous vehicle”

Neben der intermodalen Suche ermöglicht CLIP auch eine sogenannte **zero-shot Klassifikation**, das heißt die Klassifikation von Konzepten ohne explizit gelabeltes Trainingsmaterial. Aus diesem Grund wurden Änderungen zum originalen Vorhaben vorgenommen, bei dem zunächst Lexika für relevante visuelle Objekte und Konzepte erstellt werden sollten. Dies ist unter Verwendung von CLIP nicht nötig, da eine Suche nach beliebigen Objekten oder Konzepten durch eine textuelle Beschreibung möglich ist.

Zur Verbesserung der **semantischen Bildersuche mit multimodalen Informationen** haben wir mit dem entwickelten Ansatz aus AP 4 zunächst bildbeschreibende Textstellen von Patentbildern detektiert und anschließend Text- und Bild mit RoBERTa (Liu et al., 2019) und CLIP (Radford et al., 2021) extrahiert. Diese Text- und Bildmerkmale wurden in separaten Indizes gespeichert. Auf Basis einer multimodalen Anfrage bestehend aus Bild und Text wurden die Merkmale mit diesen Methoden extrahiert und zunächst einzeln mit dem dazugehörigen Index über die Kosinusähnlichkeit verglichen. Anschließend wurden die resultierenden Ähnlichkeiten aus beiden Vergleichen gemittelt, um ein neues Ranking auf Basis der multimodalen Informationen zu erstellen. Ergebnisse auf einem Subkorpus des Datenbestandes aus AP 1 mit circa 30.000 Bildern aus 2.858 Patenten und 3.770 Anfragen haben gezeigt, dass dieses multimodale Re-Ranking zu besseren Ergebnissen führt als eine Suche auf unimodalen Informationen (Bild oder Text). Diese Arbeit wurde beim Workshop *Patent Text Mining and Semantic Technologies 2021* vorgestellt (Pustu-Iren et al., 2021).

## AP 7: GUI Entwicklung / Formative Evaluation

Nachdem in den vorgelagerten Arbeitspaketen die Untersuchung des Informations- und Suchverhaltens der Patentexperten sowie die Klärung, welche Formen der Integration der automatischen Bildanalyse sich in bestehenden Infrastrukturen anwenden lassen, stattgefunden hatte, war der nächste Schritt die nutzer:innenzentrierte Entwicklung und Gestaltung eines visuellen Patentretrievalsystems. Aus der Sicht der Gestaltung war darüber hinaus der Aspekt interessant, in welchen Situationen der Einsatz der visuellen Suche der textuellen Suche vorzuziehen ist, bzw. wann und in welcher Form eine Kombination der beiden Modalitäten für spezifische Nutzerperspektiven am stärksten zielführend ist.

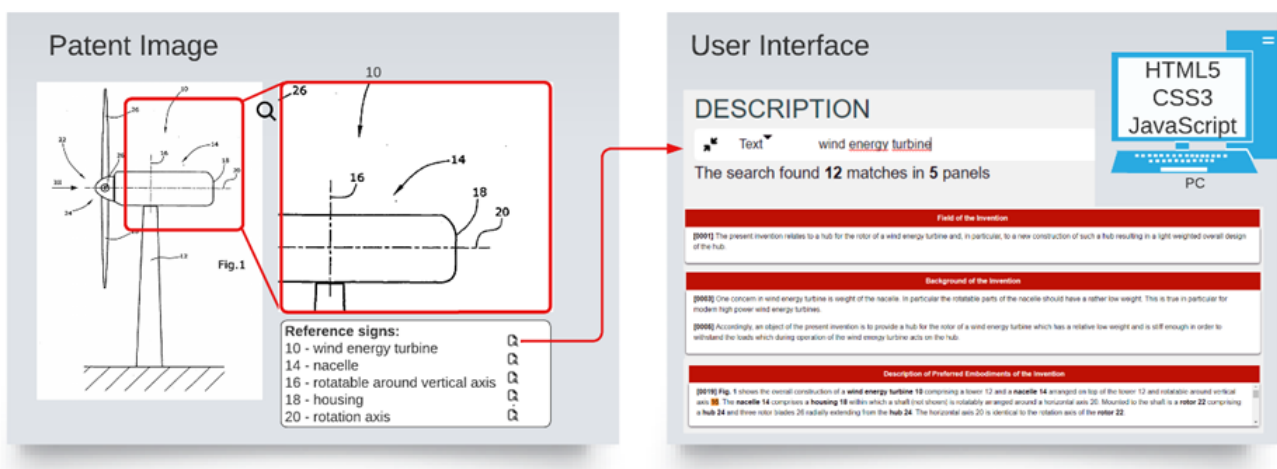
Der gesamte Entwicklungsprozess erfolgte iterativ nutzer:innenzentriert, sowohl bei der Anforderungsanalyse, als auch während der Entwicklung mittels formativer Evaluierung. Iterative Systementwicklung ist ein weit verbreiteter Ansatz für die Entwicklung komplexer Informationssysteme, der einen wiederholten Zyklus der Planung, des Entwurfs, der Entwicklung, des Testens und der Verfeinerung des Systems beinhaltet, bis es den gewünschten Grad an Funktionalität, Leistung und Benutzer:innenfreundlichkeit erreicht hat. Dabei wird ein Projekt in kleinere, besser überschaubare Phasen oder Iterationen aufgeteilt. Dieser Ansatz ermöglicht eine größere Flexibilität und Anpassungsfähigkeit bei der Reaktion auf sich ändernde Anforderungen und Rückmeldungen während des gesamten Entwicklungsprozesses. Außerdem ermöglicht er ein kontinuierliches Feedback von Interessengruppen und Nutzer:innen, was dazu beiträgt, Probleme frühzeitig im Entwicklungsprozess zu erkennen und zu beheben, wodurch das Risiko von kostspieligen Fehlern und Verzögerungen verringert wird. Der iterative Ansatz war bei der Entwicklung komplexer Informationssysteme wie dem unseren besonders vorteilhaft, da die Anforderungen und die Funktionalität zu Beginn noch nicht vollständig bekannt sind und sich während der Implementierung noch Änderungen ergeben werden.

Die Grundfunktionen wie Suche, Ergebnisdarstellung, Navigation und die persistente Datenhaltung orientieren sich grundsätzlich an Richtlinien der Usability und User Experience, sind dennoch als prototypisch anzusehen. Anforderungen sowie offene Fragestellungen sind dokumentiert und stehen einer effizienten Weiterentwicklung zur Verfügung. Die summative Evaluierung konzentrierte sich konsequenterweise auf die Kernfunktion der Bildsuche. Die Evaluierung basiert jeweils auf prototypischen Fragestellungen und ihren zugehörigen Antwortmengen, die ganzheitliche Evaluierung des Systems aus Benutzer:innenperspektive, bezogen auf die Arbeitsaufgaben der Patent-Experten, erforderte ähnliche empirische und teils qualitative Methoden wie die Informationsbedarfsanalyse.

Die zentralen Aspekte, die sich mittels der Information-Behaviour-Analyse herauskristallisierten, lagen basierend auf der Bildanalyse in der Entwicklung einer Integration von Bildinformationen (primär Referenzzeichen) in die Patentansicht sowie auf der Bild- und der semantischen Suche. Die Bildsuche funktioniert, indem ein von einer Benutzer:in hochgeladenes Bild kodiert und mit den Kodierungen der Bilder jedes einzelnen Patents in der Datenbank verglichen wird. So erhält man einen Überblick darüber, welches Patentbild dem

Suchbild am ähnlichsten ist, und kann die Patentauswahl nicht nur anhand der Patenttexte, sondern auch anhand der Bilder bewerten.

Insbesondere durch die Extraktion der Referenzzeichen aus den Bildern mittels optischer Zeichenerkennung (OCR) und der Identifikation der Referenzzeichen im Patenttext war eine interaktive Text-Bild-Integration in der Patentansicht erst möglich. Für die Nutzer:innen bedeutet das, dass direkt aus der textuellen Beschreibung eines Patents parallel das relevante Bild angezeigt wird (mittels Mouseover) sowie umgekehrt, dass aus der Bildansicht in die relevanten Textpassagen vorhandener Referenzzeichen verlinkt werden kann. Der konzeptionelle Bezug der Texte und der Bilder eines Patents erhalten so ein technisch nutzbares Äquivalent. Darüber hinaus sollte die Übernahme der Beziehungen zwischen Text und Bild in den Suchindex zu verbesserten Ergebnissen auch bei der klassischen Textsuche führen.



**Abb. 9:** Beziehung Bild-Text

Flankiert worden ist die Entwicklungsarbeit durch einen relativ hohen Anteil am Arbeitspaket „Systemintegration“ (Arbeitspaket 8). Ziel des Arbeitspakets war es, die im Rahmen des Projekts zu entwickelnden technischen Komponenten und ihre Interaktionen aus softwaretechnischer Sicht so zu gestalten, dass die funktionalen und nicht-funktionalen Anforderungen möglichst gut unterstützt werden.

Die Systembestandteile „Suchindex“, „Bildanalyse“ sowie „Grafische Benutzeroberfläche“ haben erst zu fortgeschrittener Projektlaufzeit konzeptionell und technisch zusammengefunden, wie auch bei der Entwicklung der Nutzer:innenschnittstelle sind zentrale Schnittstellendefinitionen als prototypisch zu bezeichnen, wobei es grundsätzlich relativ gut gelungen ist, die Teilsysteme technisch zu entkoppeln und somit unabhängig voneinander zu halten, was bei einer Weiterentwicklung den Austausch von Teilkomponenten erleichtern sollte. Dennoch ist die Architektur relativ heterogen, was sich in verschiedenen Programmiersprachen und unterschiedlichen technischen Ansätzen bei der Kommunikation der Teilsysteme niederschlägt.

## Werkzeuge

Zwei wesentliche Werkzeuge zur projektinternen Kommunikation waren Jira und Git. Mit Jira konnten die drei spezialisierten Entwicklungsteams - "Bildererkennung und -abfrage", "Datenbank- und Textabfragetechniken" und "GUI-Design" - Aufgaben erstellen und verwalten, den Fortschritt verfolgen und Verantwortlichkeiten für jede Iteration zuweisen. Git wiederum half den Entwickler:innen dabei, Codeänderungen zu verwalten, die Versionskontrolle aufrechtzuerhalten und miteinander zusammenzuarbeiten. Eine weitere wichtige Methode des iterativen Systementwicklungsprozesses war die Einrichtung wöchentlicher virtueller Besprechungen ("Jour Fixes"). Diese boten den Teammitgliedern die Möglichkeit, Fortschritte zu abzustimmen, Hindernisse zu erkennen und operativ die Weiterentwicklung zu planen.

## Systemkomponenten

Angesichts der Komplexität von Patentrecherche-Systemen erwies es sich als undurchführbar, einfach ein visuelles Recherchemodul zu entwickeln, das leicht in bestehende Systeme integriert werden kann. So kam es zum Entschluss, den Prototyp eines eigenen vollwertigen Patentrecherchesystems zu entwickeln (VisPat). Als solches kombiniert es Bildsuchfunktionen mit lexikalischen und semantischen Suchoptionen und ermöglicht so Text-Bild-Referenzen. Das System bietet außerdem Filteroptionen zur Verfeinerung der Suchergebnisse.

VisPat besteht aus verschiedenen Komponenten und greift derzeit auf eine Entwicklungsdatenbank mit rund 34.000 Patenten aus den Technologiebereichen Windenergie und autonomes Fahren zurück. Die Patente liegen als XML- und PDF-Dokumente und die Patentbilder als TIF-Dateien vor. In der Vorverarbeitungsphase zerlegt der XML-Parser die XML-Dateien in fünf Abschnitte: BIBLIOGRAPHISCHE DATEN, BESCHREIBUNG, ANSPRÜCHE, ZEICHNUNGEN und ZITATE. Die Inhalte und die zugehörigen Metadaten werden in einer SQL-Datenbank gespeichert, um einen effizienten Zugriff und eine weitere Vorverarbeitung zu ermöglichen und die Benutzer:innenoberfläche zu gestalten.

## AP 11: Projektmanagement

Management und Koordination des Projekts wurden im Wesentlichen von der TIB geleistet, einzelne Teile von IAIS und LG. Die Stelle wurde von einer bereits in der TIB beschäftigten Mitarbeiterin besetzt. In der zweiten Jahreshälfte 2022 wurden besonders Aufgaben im IT Management von einem zusätzlichen Mitarbeiter übernommen. Dieser übernahm ab dem 1. Januar 2023 gemeinsam mit einem Mitarbeiter der Geschäftsstelle der Leibniz-Gemeinschaft auch die Aufgaben des Projektmanagements, da die bisherige Projektleiterin aus dem Projekt ausschied.

Zur Projektkoordinierung wurden zu Beginn eine Mailingliste beim DFN und ein Space im Confluence-Wiki der TIB eingerichtet. Das Projektmanagement koordinierte und plante die monatlichen Projekttreffen und übernahm Protokoll und Nachbereitung. Protokolle aller Treffen

wurden ins Wiki gestellt und dort nachhaltig archiviert. Auch die wöchentlichen Arbeitstreffen der Entwickler:innen wurden vom PM organisiert und im Wiki protokolliert. In der Endphase des Projekts wurden die Aufgaben im IT-Management intensiviert, um die Entwicklungsprozesse zu optimieren. Das PM organisierte und protokollierte die wöchentlichen Developer Meetings, strukturierte die Arbeit in Jira neu, organisierte den Workflow in der TIB eigenen GitLab Instanz und setzte wöchentliche Ziele und Deadlines. Die unterschiedlichen Teile der Webapplikation wurden mithilfe eines Miroboards visualisiert und schließlich mithilfe diverser Technologien und Frameworks integriert. Diese Integration wurde in der Endphase des Projekts ebenfalls vom TIB Projektmanagement koordiniert. Daneben sorgte das PM für die Dokumentation der von den unterschiedlichen Partnern verantworteten Entwicklungen.

Weitere Tätigkeiten des PM umfassten die Organisation von drei größeren internen Workshops (06.07.2021, 05.10.2021, 05.05.2023), die Erstellung von Entwicklungs-Roadmaps und die diesbezügliche Zielkontrolle sowie die externe und TIB-interne Kommunikation, u.a. auch durch einen Newsletter. Desweiteren oblag dem PM die Kommunikation mit dem Projektbeirat sowie die Organisation zweier Beirats-Treffen (08.10.2021, 02.06.2023).

In Zusammenarbeit mit der LG wurde ein Logo entwickelt und ein inhaltliches Konzept für ein kurzes Video zur Einführung des Anwendungsprototyps erarbeitet. Das Video wurde schließlich von Mitarbeitern des *Lab für Nicht-Textuelle Materialien* der TIB produziert. Dem Projektmanagement oblagen auch die allgemeinen formalen Aufgaben wie Mittelverwaltung (siehe dazu 2.2), interne Dokumentation -- organisiert über ein Confluence Wiki -- sowie die Berichterstattung an den Projektträger. Projektspezifisch war die Koordinierung des Unterauftrags an das IWIST (Uni Hildesheim), mit dem dazu regelmäßige Treffen stattfanden.

## 2.2 Mittelverwendung

Der TIB standen insgesamt Projektmittel in Höhe von 462.912,54€ zur Verfügung, von denen 454.240,33€ verausgabt wurden. Die Differenz erklärt sich aus den aufgrund der Coronakrise kaum in Anspruch genommenen Mitteln für Dienstreisen (Posten 0846). Die Mittel wurden im Wesentlichen für Personalkosten (240.313,63€, Posten 0812) sowie den externen Auftrag an das IWIST (189.642,26€, Posten 0835) genutzt. Daneben fielen Verwaltungsausgaben in Höhe von 24.031,36€ an. Die Personalkosten der TIB entsprachen einer TV-L 13 Stelle, die sich zu 3/4 auf eine/n wissenschaftliche/n Mitarbeiter:in für die Software-Entwicklung und zu 1/4 für die Projektkoordination verteilte. Die Aufgaben wurden während der Projektlaufzeit von mehreren Personen wahrgenommen.

## 2.3 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Wesentliche Arbeiten des gesamten Projekts wie auch des entwickelten Prototypen wurden von der TIB geleistet. Speziell die Entwicklung der Methoden und Algorithmen zur Bildanalyse stellten ein Kernelement der geleisteten Entwicklungs- und Forschungsarbeit dar, deren Ergebnisse auch für weitere Projekte nicht nur in der Patentdomäne nutzbar sein werden. Die

Entwicklung des GUI auf Basis einer extensiven Nutzer:innenanalyse im Feld der Patentexpert:innen sowie die Zusammenführung der einzelnen Programmteile (Bildanalyse, Suchfunktionen, GUI, Webframework) in einer Webapplikation wurden von der TIB, bzw. vom Unterauftragnehmer IWIST, geleistet und koordiniert.

Das Projektmanagement führte die üblicherweise notwendigen Arbeiten durch, war darüberhinaus aber mit der Corona-Krise und der hohen Personalfuktuation vor besondere Herausforderungen gestellt. Im letzten Projektdrittel leistete das Management zusätzlich erhebliche Arbeiten in der auch technischen Koordinierung der Entwicklungsarbeit.

## 2.4 Nutzen und Verwertbarkeit

Im Rahmen des Projekts ExpResViP wurde ein Prototyp zur bildbasierten Suche in Patenten entwickelt. Diese Funktionalität stellt ein Alleinstellungsmerkmal gegenüber existierenden Patentsuchsystemen (PIS) dar, die sich vorrangig auf die Suche mithilfe von Metadaten oder Textinformationen beschränken. Damit hat das Projekt erfolgreich dazu beigetragen, PIS in Zukunft zu verbessern. Der entwickelte Prototyp *VisPat* zur bildbasierten Patentsuche wurde auf einem Korpus bestehend aus zwei Patentdomänen (autonomes Fahren, Wind- und Wasserenergie) eingesetzt und evaluiert. Der Webservice wurde der interessierten Community zur Verfügung gestellt.

Ein Teilvorhaben der TIB fokussierte sich dabei auf die Entwicklung von grundlegenden und weiterführenden Algorithmen zur Patentbildanalyse. Der erstellte Datensatz zur Klassifikation von Abbildungsformen und Bildperspektiven ist inklusive Quellcode für die Nachnutzung veröffentlicht (Ghauri et al., 2023), selbiges gilt für weitere Teile der Applikation. An der TIB (Forschungsgruppe Visual Analytics) sind Themen für studentische Abschlussarbeiten ausgeschrieben, um die Entwicklung des Systems weiter voranzutreiben. Ein Fokus liegt auf der Optimierung von CLIP für die Patentdomäne und der Erschließung weiterer Patentklassen. Verwertungspotenziale für künftige Forschungen und Entwicklungen liegen darüberhinaus auch in der intensiver als im Projektrahmen möglich weiteren Auswertung und Rückkoppelung der Nutzer:inneninterviews.

Die Projektergebnisse fließen in aktuelle Forschungsvorhaben ein, die im Laufe des Projekts ExpResViP begannen, etwa im Rahmen des Projektes ViP@Scale welches durch das europäische Patentamt gefördert wird. Die Einreichung von Folgeanträgen bei nationalen und internationalen Ausschreibungen wird geprüft und könnte etwa auf der Erweiterung der Plattform mithilfe von Methoden der generativen und erklärbaren KI liegen.

## 2.5 Projektrelevante Fortschritte Dritter

Eine zur Zeit des Projektantrags noch nicht existierende Forschungsrichtung bestand in Vision-Language-Modellen wie CLIP (Contrastive Language-Image Pretraining; Radford et al., 2021). CLIP wurde mit 400 Millionen Bild-Text-Paaren trainiert und ermöglicht die Klassifikation von beliebigen Konzepten über textuelle Beschreibungen. Durch die hohe Anzahl an verwendetem Trainingsmaterial ist die Güte bei der Erkennung von Objekten selbst ohne die Verwendung von zusätzlichen spezifischen Materialien sehr gut. Aus diesem Grund haben wir uns im Laufe des Projektes entschieden, CLIP für die bildbasierte Ähnlichkeitssuche und zur Erkennung von Objekten (AP 5) einzusetzen. Durch die cross-modale Suche über beliebige Benutzer:innendefinierte Beschreibungen, konnte auf eine explizite Definition eines Konzeptlexikons verzichtet werden.

Weiterhin sind in den letzten Jahren Modelle wie CLIP mit Methoden der generativen AI kombiniert worden (z.B. Li et al. 2023). Aubakirova et al. 2023 haben gezeigt, wie mithilfe solcher Modelle beispielsweise automatisch Bildunterschriften für Abbildungen in Patenten generiert werden können, die sowohl zur Patentsuche aber auch zum Training von Vision-Language-Modellen eingesetzt werden können. Eine weitere Chance besteht darin, dass Nutzer:innen interaktiv beispielsweise eine Facettierung der Ergebnisse durchführen oder Fragen für eine erhöhte Erklärbarkeit der Ergebnisse stellen können. Diese Forschungsrichtung soll in einem gemeinsamen Folgeantrag weiter verfolgt werden.

## 2.6 Publikationen

Ghuri, J. A., Müller-Budack, E., & Ewerth, R. (2023). Classification of Visualization Types and Perspectives in Patents. In International Conference on Theory and Practice of Digital Libraries, TPD 2023, pp. 182-191. Cham: Springer Nature Switzerland.

Pustu-Iren, K., Bruns, G., & Ewerth, R. (2021). A multimodal approach for semantic patent image retrieval. In Patent Text Mining and Semantic Technologies co-located with the International ACM SIGIR Conference on Research and Development in Information Retrieval, PatentSemTech@SIGIR 2021. pp 45-49.

Zellmer, J., Elbeshausen, S., & Womser-Hacker, C. (2022). Elicitation of requirements for innovative visual patent retrieval based on interviews with experts. In Proceedings of ISIC: the information behaviour conference, Berlin, Germany, 26-29 September, 2022. Information Research, 27 (Special issue), isic2234. <https://doi.org/10.47989/irisic2234>

## 3. Erfolgskontrollbericht

### 3.1 Wissenschaftlich-technisches Ergebnis

Die Ergebnisse des Projekts sind prototypisch in der Webapplikation *VisPat* manifestiert.<sup>8</sup> *VisPat* kann als Patentrecherchesystem genutzt werden. Es verbindet dabei innovative Bildsuchfunktionen sowie eine lexikalische und semantische Suche, ermöglicht die Suche nach Text-Bild-Referenzen und bietet Filteroptionen. Die Datenbank enthält derzeit 34.000 Patente aus den Domänen *Windenergie* und *autonomes Fahren*. Die Applikation dient als *Proof of Concept* und bietet hierfür eine einfache, aber funktionale Oberfläche. Sie besteht neben der Patentdatenbank aus verschiedenen containerbasierten Komponenten für Bildanalyse, Suche und Userinterface.

Wissenschaftlich weiter nutzbare Ergebnisse sind die entwickelten Methoden und Algorithmen zur patentspezifischen Bildanalyse sowie die Klassifikation von Abbildungsformen und Perspektiven, die bereits publiziert wurden (Ghuri et al., 2023). Im Bereich Usability konnten Erkenntnisse zum Informations- und Suchverhalten von Patentexpert:innen gewonnen sowie daraus abgeleitet Anforderungen an eine effektive bildbasierte Patentrecherchelösung formuliert werden.

### 3.2 Fortschreibung des Verwertungsplans

Am Markt gibt es nach im Rahmen des Projekts von der Leibniz-Gemeinschaft durchgeführten Recherchen bisher kein Patentinformationssystem, das den vollen Informationsgehalt von Bildelementen in Patentdokumenten und dessen Verknüpfung mit textlichen Information ausnutzt. Der Ansatz des Projekts kann daher weiterhin als innovativ und seine allgemeinen wirtschaftlichen Verwertungschancen grundsätzlich als günstig gelten (angesichts des aktuellen rasanten Fortschritts in den relevanten Technologiebereichen ist diese Bewertung jedoch unter einem starken zeitlichen Vorbehalt zu betrachten). Nach Einschätzung der Projektpartner gibt es im Bereich der Patentrecherche nach wie vor ein großes wirtschaftliches Potenzial für Lösungen, die das Informationsbedürfnis gerade auch von Nutzer:innen, die über keine starke Patentexpertise verfügen, effektiver und effizienter zu bedienen als bestehende Angebote.

Voraussetzung für die weitere Verwertbarkeit wäre allerdings eine auf den bisherigen Ergebnissen des Projekts aufbauende sowie neuere Forschungen und Modelle wie CLIP aufnehmende Verbesserung der Bildanalyse sowie eine Weiterentwicklung des Prototypen *VisPat* zu einer produktiv nutzbaren Applikation. Die TIB selbst verfolgt keine wirtschaftlichen Interessen und plant keine kommerzielle Nutzung der Ergebnisse. Die von der TIB entwickelten

---

<sup>8</sup> <https://service.tib.eu/vispat>

Verfahren und Softwarekomponenten werden veröffentlicht und stehen zur weiteren Nutzung und Verarbeitung, wissenschaftlicher wie kommerzieller, zur Verfügung.

Die Projektpartner betrachten die bisherigen Ergebnisse als vielversprechende Grundlage für weitere gemeinsame Vorhaben. Das Konsortium plant einen Folgeantrag, für den eine Skizze bereits beim BMBF vorgelegt wurde und in dem vor allem Einsatz und Entwicklung neuartiger Methoden im Bereich der generativen und erklärbaren KI stehen sollen.

### 3.3 Arbeiten, die zu keiner Lösung geführt haben

Alle im Projekt durchgeführten Arbeiten haben zu Lösungen für die jeweilige Aufgabenstellung geführt.

### 3.4 Einhaltung der Ausgaben- und Zeitplanung

Die Zeitplanung des Projekts musste mehrfach angepasst und modifiziert werden. Wesentlicher Faktor war hier zunächst die Corona-Krise, die eine Umstellung der Kooperationsformen erforderte, zu Einschränkungen bei der Durchführung von Workshops und Interviews führte und Stellenbesetzungen verzögerte. In der Folgezeit kam es bei allen Projektpartnern weiter zu Personalfluktuationen. Bei der TIB bezog sich das auf die Entwickler:innen-Stelle sowie ein gutes halbes Jahr vor Projektende die Koordinationsstelle.

Trotz dieser Schwierigkeiten konnten wesentliche Ziele des Projekts im Projektzeitraum erreicht werden. Entwicklungsseitig wurden Rückstände im letzten Projektjahr aufgeholt. Die entstandene Vakanz der Projektkoordination wurden durch einen TIB-internen Mitarbeiter sowie einen Mitarbeiter der Leibniz-Gemeinschaft aufgefangen.

Die geplanten Ausgaben wurden nicht vollständig in Anspruch genommen, was im Wesentlichen auf die weitgehend entfallenen Reise- und Sachkosten zurückzuführen ist.

# Referenzen

Aubakirova, D., Gerdes, K., & Liu, L. (2023). PatFig: Generating Short and Long Captions for Patent Figures. In IEEE/CVF International Conference on Computer Vision, ICCV 2019, pp. 2843-2849.

Azzopardi, Leif , Vanderbauwhede, Wim ; Joho, Hideo ( 2010). Search system requirements of patent analysts. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 775-776).

Baek, Y., Lee, B., Han, D., Yun, S., & Lee, H. (2019). Character region awareness for text detection. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, pp. 9365-9374. <https://api.semanticscholar.org/CorpusID:102480461>

Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., Oh, S.J. and Lee, H., (2019). What is wrong with scene text recognition model comparisons? dataset and model analysis. In the International Conference on Computer Vision, ICCV 2019, pp. 4715-4723. <https://api.semanticscholar.org/CorpusID:102481180>

Kembhavi, A., Salvato, M., Kolve, E., Seo, M. J., Hajishirzi, H., & Farhadi, A. (2016). A Diagram is Worth a Dozen Images. In European Conference on Computer Vision, ECCV 2016, pp. 235-251. <https://api.semanticscholar.org/CorpusID:2682274>

Kuckartz, Uwe (2018). Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Beltz Juventa, Weinheim.

Li, J., Li, D., Savarese, S., & Hoi, S. (2023). Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. arXiv preprint arXiv:2301.12597.

Liao, M., Wan, Z., Yao, C., Chen, K. and Bai, X., (2020), April. Real-time scene text detection with differentiable binarization. In the Conference on Artificial Intelligence, AAAI 2020, Vol. 34, No. 07, pp. 11474-11481. <https://api.semanticscholar.org/CorpusID:208175645>

Liu, W., Chen, C., Wong, K.-Y. K., Su, Z., & Han, J. (2016). STAR-Net: A SpaTial Attention Residue Network for Scene Text Recognition. In British Machine Vision Conference 2016, BMVC 2016, pp. 19-22. <https://api.semanticscholar.org/CorpusID:22482128>

Long, S., Ruan, J., Zhang, W., He, X., Wu, W. and Yao, C. (2018). Textsnake: A flexible representation for detecting text of arbitrary shapes. In European Conference on Computer Vision, ECCV 2018, pp. 20-36. <https://api.semanticscholar.org/CorpusID:49570059>

Mayring, Philipp (2015). Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12. überarbeitete Aufl. Beltz, Weinheim, Basel.

Piroi, F., Lupu, M., Hanbury, A., & Zenz, V (2011). CLEF-IP 2011: Retrieval in the intellectual property domain. In: CLEF 2011 Labs and Workshop, Notebook Papers, 19-22 September 2011, Amsterdam, The Netherlands. CEUR Workshop Proceedings, vol. 1177. CEUR-WS.org (2011), <https://ceur-ws.org/Vol-1177/CLEF2011wn-CLEF-IP-PiroiEt2011.pdf>

Piroi, F., & Hanbury, A. (2017). Evaluating Information Retrieval Systems on European Patent Data: The CLEF-IP Campaign. In Current Challenges in Patent Information Retrieval, The Information Retrieval Series (Vol. 37), pp. 113-142. Springer. [https://doi.org/10.1007/978-3-662-53817-3\\_4](https://doi.org/10.1007/978-3-662-53817-3_4)

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. In International Conference on Machine Learning, ICML 2021, pp. 8748–8763. PMLR. <http://proceedings.mlr.press/v139/radford21a.html>

Shalaby, Walid , Zadrozny, Wlodek (2019). Patent retrieval: a literature review. Knowledge and Information Systems, 61,5-6. DOI: <https://doi.org/10.1007/s10115-018-1322-7>

Wei, X., Wu, J., Ajayi, K., & Oyen, D. (2022). Visual descriptor extraction from patent figure captions: A case study of data efficiency between BiLSTM and transformer. In ACM/IEEE Joint Conference on Digital Libraries, JCDL 2022, pp. 1-5. <https://doi.org/10.1145/3529372.3533299>

Xie, S., Girshick, R. B., Dollár, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 5987–5995. IEEE. <https://doi.org/10.1109/CVPR.2017.634>

# Anlagen

Anlage 1: Leitfragen Expert:inneninterviews

## Projekt ExpResViP

- Fokus:
  - Bedeutung von Bildern in Patentschriften für die Suche
  - Bezug zwischen Bild und Text → Bedeutung für Suche

## Leitfragen

- 1. Bitte erzählen Sie zunächst etwas über Ihre Tätigkeit und Ihren beruflichen Hintergrund.**
  - a. In welchem Bereich arbeiten Sie?
  - b. Was ist Ihr Fachgebiet?
  - c. Mit welcher Art von Patenten befassen Sie sich?
  - d. Welchen Anteil haben Patente und Patentsuchen in Ihrem Arbeitsalltag?
- 2. Bitte schildern Sie, welche Art von Patentrecherchen Sie durchführen und wie Sie dabei genau vorgehen.**
  - a. Was sind die Rechercheanlässe für die Suchen?
  - b. Welche Ressourcen verwenden Sie?
  - c. Welche/-s System/-e verwenden Sie für die Suche?
  - d. Welche Art von Information ist für Sie relevant? Nur das Patent selbst? Ziehen Sie auch andere Quellen heran?
  - e. Welche Probleme können (typischerweise) während der Suche auftreten?
  - f. Erinnern Sie sich vielleicht an eine konkrete Situation, die problematisch war und können diese kurz beschreiben?
- 3. Welche Rolle spielen Patentzeichnungen während der Suche für Sie?**
  - a. Nutzen Sie diese?
  - b. Wenn ja, können Sie bitte beschreiben, wie Sie Bilder in Ihre Arbeit einbinden?
  - c. Vielleicht erinnern Sie sich an eine konkrete Situation, in der Sie gezielt nach Zeichnungen gesucht haben oder die Patentzeichnung für Ihre Recherche von besonderer Relevanz war?
- 4. Welchen Stellenwert hat der Text im Patentdokument für Ihre Suche?**
  - a. Welche Textabschnitte verwenden Sie am häufigsten und aus welchen Gründen?
- 5. In Patenten gibt es in der Regel Textabschnitte, welche die Zeichnungen erläutern. Wo ist der Bezug zwischen Text und Bild am größten? Wie schätzen Sie die inhaltliche Bedeutung dieser Abschnitte ein?**
  - a. Nutzen Sie diese Textabschnitte für Ihre Recherche?
  - b. Welche Bedeutung haben diese Textabschnitte für Ihren Rechercheprozess?
- 6. Können Sie eine Einschätzung geben, welche Anteile Patentzeichnungen, Patenttexte und die erläuternden Textabschnitte zu den Zeichnungen für Ihre Suche haben?**
- 7. Gibt es bestimmte Situationen, in denen Sie Texte, Zeichnungen oder textuelle Erläuterungen zu den Zeichnungen vorziehen?**

- a. Gibt es spezifische Rechercheanlässe oder Problemlagen, in denen Zeichnungen besonders relevant sind?
- b. Gibt es spezifische Rechercheanlässe, in denen Text besonders relevant sind?
- c. Gibt es spezifische Rechercheanlässe, in denen die Kombination aus Text und Zeichnungen besonders relevant ist?

**8. Gibt es bestimmte Situationen, in denen Sie eine visuelle Suche vorziehen würden, es aber nicht möglich ist?**

- a. Wenn ja, was sind die Hinderungsgründe?
- b. Wenn ja, was würde Ihnen ggf. helfen oder die visuelle Recherche vereinfachen?

**9. Text und Bild**

- a. Wie nützlich würden Sie eine Funktion einschätzen, die bestimmte Objekte oder Texte innerhalb von Abbildungen sucht? (z.B. auch Referenzzahlen)
- b. Sind Beziehungen zwischen dem Text innerhalb von Abbildungen und dem Text im Dokument von Interesse für Sie?
- c. In welchen Situationen und wofür wäre dies von Vorteil?

**10. Gibt es noch etwas, was Sie hinzufügen möchten?**

## Anlage 2: Kategorienbildung

Kategorie	Unterkategorie	Definition
K1: Beruflicher Hintergrund	U1.1: Tätigkeitsbereich	Aktuelle Berufliche Tätigkeit sowie Hintergrund (Studium etc.)
	U1.2: Bezug zu Patenten	Anteil von Patenten und Patentsuchen im beruflichen Alltag (ggf. auch Patentkategorie)
K2: Vorgehen Patentrecherche	U2.1: Rechercheanlässe im beruflichen Alltag	Nennung von Rechercheanlässen wie Neuheit/Patentierbarkeit, Stand der Technik, Gültigkeit/Opposition, Freedom to Operate/Ausübungsfreiheit
	U2.2: Ressourcen, Systeme, Quellen	Alle genutzten Systeme und Ressourcen; Patente selbst oder auch Nicht-Patentliteratur als Quellen
	U2.3: Probleme Patentrecherche	Schwierigkeiten, die bei der Patentrecherche typischerweise, also regelmäßiger, auftreten. Ggf. konkrete Situationen.
K3: Rolle und Einbindung von Patentzeichnungen	U3.1: Nutzung von Patentzeichnungen U3.1.1: Nutzung positiv U3.1.2: Nutzung negativ	U3.1.1: Patentzeichnungen werden im Rechercheprozess genutzt. U3.1.2: Patentzeichnungen werden im Rechercheprozess nicht genutzt.
	U3.2: Begründungen für (Nicht-)Nutzung U3.2.1: Gründe für die Nutzung U3.2.2: Gründe gegen die Nutzung	U3.2.1: <b>Warum</b> werden die Zeichnungen bei der Recherche genutzt? Nennung von Gründen (bspw. Rechercheanlässe, Patentkategorien), warum gezielt nach Zeichnungen gesucht wird oder Zeichnungen insgesamt sehr hilfreich sind und einen hohen Stellenwert haben. Ggf. allgemeine Vorteile von Zeichnungen. U3.2.2: <b>Warum</b> werden die Zeichnungen bei der Recherche nicht genutzt? Nennung von Gründen (bspw. Rechercheanlässe, Patentkategorien), warum Zeichnungen keine Rolle spielen und ggf. sogar vom Rechercheprozess ausgeschlossen werden. Ggf. allgemeine Nachteile von Zeichnungen.
	U3.3: Einbindung von Patentzeichnungen in die Recherche	<b>Wie</b> werden die Zeichnungen im Rechercheprozess eingebunden? <b>Wann</b> werden sie eingebunden bzw. an welcher Stelle im Rechercheprozess sind sie am relevantesten?
atenttexten	U4.1: Rolle von Text	Äußerungen über den Stellenwert von Patenttext im Rechercheprozess.

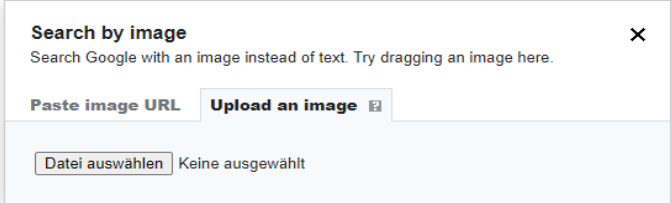
K4: Stellenwert von P	U4.2: Einbindung von von Patenttexten in die Recherche	<b>Wie</b> werden Patenttexte im Rechercheprozess eingebunden? <b>Wann</b> werden sie eingebunden bzw. an welcher Stelle im Rechercheprozess sind sie am relevantesten?
K5: Rolle und Nutzung von erläuternden Textabschnitten für Patentzeichnungen	U5.1: Nutzung von erläuterndem Text U5.1.1: Nutzung positiv U5.1.2: Nutzung negativ	U5.1.1: Erläuternder Text wird im Rechercheprozess genutzt. U5.1.2: Erläuternder Text wird im Rechercheprozess nicht genutzt.
	U5.2: Begründungen für (Nicht-)Nutzung U5.2.1: Gründe für die Nutzung U5.2.2: Gründe gegen die Nutzung	U5.2.1: <b>Warum</b> wird der erläuternde Text bei der Recherche genutzt? Nennung von Gründen (bspw. Rechercheanlässe, Patentkategorien), warum der erläuternde Text sehr hilfreich und von hoher Relevanz ist. U5.2.2: <b>Warum</b> wird der erläuternde Text bei der Recherche nicht genutzt? Nennung von Gründen (bspw. Rechercheanlässe, Patentkategorien), warum der erläuternde Text keine Rolle spielt und ggf. sogar vom Rechercheprozess ausgeschlossen wird.
	U5.3: Einbindung von erläuterndem Text in die Recherche	<b>Wie</b> wird der erläuternde Text im Rechercheprozess eingebunden? <b>Wann</b> wird er eingebunden bzw. an welcher Stelle im Rechercheprozess ist er am relevantesten?
K6: Möglichkeit visueller Suche	U6.1: Wunsch nach visueller Suche U6.1.1: Wunsch vorhanden U6.1.2: Wunsch nicht vorhanden	U6.1.1: Äußerungen darüber, dass eine visuelle Suche in bestimmten Situationen gerne vorgezogen würde. U6.1.2: Äußerungen darüber, dass eine visuelle Suche in keiner Situation gewünscht war.
	U6.2: Situationen, in denen visuelle Suche gewünscht war	Nennung von Situationen, in denen der Wunsch nach visueller Suche vorhanden war und vorgezogen würde. Nennung von Gründen.
	U6.3: Hinderungsgründe für visuelle Suche	Äußerungen darüber, warum die visuelle Suche nicht durchgeführt werden konnte. Bezug zu bspw. Problemen seitens der Systeme oder der Beschaffenheit der Patente selbst.
	U6.4: Vorschläge für Erleichterung visueller Suche	Ideen für die Erleichterung der visuellen Suche; ggf. Vorschläge ideales System

K7: Beziehungen zwischen Text und Zeichnung	<p>U7.1: Bedeutung von Beziehungen zwischen Texten und Zeichnung  U7.1.1: Beziehung ist wichtig/interessant  U7.1.2: Beziehung ist unwichtig/uninteressant</p>	<p>U7.1.1: Äußerungen darüber, dass die Beziehungen zwischen den Patenttexten und den Zeichnungen, speziell einzelne Bauteile oder Bezugszeichen, interessant und bedeutend sind. Äußerungen über die (inhaltliche) Bedeutung eben dieser Beziehung.</p> <p>U7.1.2: Äußerungen darüber, dass die Beziehungen zwischen den Patenttexten und den Zeichnungen, nicht interessant oder unbedeutend sind.</p>
	<p>U7.2: Situationen, in denen Beziehung zwischen Text und Zeichnung eine Rolle spielen</p>	<p>Nennung von Situationen, in denen die Beziehung zwischen Text und Zeichnung eine Rolle spielen sowie Nennung von Gründen.</p>
	<p>U7.3: Einschätzung Nützlichkeit von Funktion für Suche von Bereichen innerhalb von Zeichnungen  U7.3.1: Funktion ist nützlich  U7.3.2: Funktion ist nicht nützlich</p>	<p>U7.2.1: Äußerungen darüber, dass eine Funktion, die bestimmte Bereiche wie Bauteile oder Bezugszeichen in Zeichnungen suchen kann, nützlich ist.</p> <p>U7.2.2: Äußerungen darüber, dass eine Funktion, die bestimmte Bereiche wie Bauteile oder Bezugszeichen in Abbildungen suchen kann, nicht nützlich ist.</p>
	<p>U7.4: Situationen, in denen eine Funktion besonders nützlich ist</p>	<p>Nennung von Situationen, in denen eine Funktion, die bestimmte Bereiche wie Bauteile oder Bezugszeichen in Zeichnungen suchen kann, nützlich ist sowie Nennung von Gründen.</p>

## Anlage 3: Requirements

## Requirements extended

### 1) Filtering/searching for information

ID	Requirement	Comments
R1	<p>The system must offer users the possibility of performing a search query with images only as well as in combination with keywords, IPCs, or other metadata.</p> <p><b>Description:</b> It must be possible to search with one or several images (e.g., images from previous searches, excerpts from images). Accordingly, the system must provide the upload of an image (e.g., as in Google image search, see figure below).</p>  <p>Furthermore, it must be possible to combine the image search with a textual search and parameters, i.e. keywords/phrases and parameters such as classes, parts of the patent etc. must be able to be added to the image (cf. R3 and R4).</p> <p>The idea to this requirement is that users can take steps of their own choosing, e.g.:</p> <p>Start with image search → result set too large, therefore restriction with parameters → if necessary, restriction again with further/other parameters</p>	
R2	<p>The system must offer users (automatic) keyword expansion.</p> <p><b>Description:</b> In order to identify similar keywords, the system must be linked to a keyword expansion tool, so that users are also suggested synonyms, meronyms and other related terms.</p>	
R3	<p>The system must offer users the possibility to filter patents.</p> <p><b>Description:</b> There must be a possibility to filter patents using parameters, e.g.:</p> <ul style="list-style-type: none"> <li>- Inventors</li> <li>- Applicants</li> <li>- Patent class</li> <li>- Patent number</li> <li>- Patent family</li> </ul>	

	<ul style="list-style-type: none"> <li>- Disclosure year</li> <li>- Legal status (e.g., all granted patents)</li> <li>- Date of publication</li> <li>- Parts of the patent (e.g., title, abstract, claims, description of embodiments, etc., but also combinations like drawing and abstract)</li> <li>- language</li> </ul> <p>In addition, it must be possible to link any parameter with keywords in the sense of "mix &amp; match".</p>	
<b>R4</b>	<p>The system must offer users the possibility to filter the drawings.</p> <p><b>Description:</b> There must be a possibility to filter drawings using parameters, e.g:</p> <ul style="list-style-type: none"> <li>- Inventors</li> <li>- Applicants</li> <li>- Patent class</li> <li>- Patent number</li> <li>- Patent family</li> <li>- Disclosure year</li> <li>- Legal status (e.g., all granted patents)</li> <li>- Date of publication</li> <li>- Language of the related patent</li> </ul> <p>In addition, it must be possible to filter according to the metadata of the drawings. Metadata are e.g.:</p> <ul style="list-style-type: none"> <li>- Type of drawing (technical drawing, process, electrical circuit, etc.)</li> <li>- Included components in combination with their reference signs</li> <li>- Sectional plane / perspective</li> <li>- Outline shape (circular, rectangular, etc.)</li> </ul> <p>Furthermore, it must be possible to link any parameters/metadata with keywords in the sense of "mix &amp; match".</p>	

## 2) Displaying/viewing information

<b>ID</b>	<b>Requirement</b>
<b>R5</b>	<p>The system must allow users to view different parts of the patent side by side.</p> <p><b>Description:</b> Users must be provided with the view of certain parts of the patent in the form of "window tiling". There must be a possibility to display different windows for e.g., the description, the drawings, the claims side by side. In particular, it should be possible to display the list of references next to the drawings.</p>

<p><b>R6</b></p>	<p>The system must allow users to select different views for the drawings.</p> <p><b>Description:</b> Users should be able to view</p> <ul style="list-style-type: none"> <li>- all drawings of a patent application at once,</li> <li>- the summary drawing (must be automatically recognized as such) or other drawings in single view with all details,</li> <li>- two or more drawings side by side, <ul style="list-style-type: none"> <li>o from one application</li> <li>o from different applications</li> </ul> </li> </ul> <p>"Window Tiling" (see also R5) can be used to allow a direct comparison between drawings.</p>
<p><b>R7</b></p>	<p>The system must offer users the possibility to scroll and move drawings.</p> <p><b>Description:</b> Drawings within an overview must be scrollable. In addition, it must be possible to zoom and rotate the drawings within a single view.</p>
<p><b>R8</b></p>	<p>The system must allow users to highlight the reference signs in the text in color.</p> <p><b>Description:</b> It must be possible to display all reference signs in the text in the form of a color highlighting. Here, the reference signs for one and the same component are to be highlighted in the same color. In addition, a scroll bar could display the different colors, so that it is clear where which reference signs occur (e.g., most frequently).</p>
<p><b>R9</b></p>	<p>The system must offer users the option of highlighting the individual components in the drawings in color.</p> <p><b>Description:</b> Within a drawing, it must be possible to highlight the individual components by highlighting the corresponding reference signs in color and, if possible, the guide lines and border lines of the components.</p>
<p><b>R10</b></p>	<p>The system must enable user to jump „on click“ from one specific component to the corresponding reference sign in the text and vice versa.</p> <p><b>Description:</b> By "on click" on the reference signs in the text, the users should be able to jump to the corresponding component in the drawing, which is then highlighted.</p> <p>In order to jump to a specific explanatory textual passage by „on click“ on a reference sign in the drawing, two navigational pathways / jump labels must be taken into account:</p> <ul style="list-style-type: none"> <li>- If the users had previously jumped from a reference sign in the text to the corresponding reference sign in the drawing, the view jumps „on click“ back to the reference sign's textual explanation. The reference sign is highlighted.</li> <li>- If the users had not previously jumped from a reference sign in the text to the corresponding reference sign in the drawing, the view jumps „on click“ to the first textual passage where the reference sign appears. The reference sign is highlighted.</li> </ul>

## 2a) Intuitive user navigation and involvement

<b>ID</b>	<b>Requirement</b>
<b>R11</b>	<p>The system must offer the users possible further search steps.</p> <p><b>Description:</b> Suggestions for further search steps can guide users through the search process and support them in this. Potential search steps can be realized for example</p> <ul style="list-style-type: none"> <li>- by suggestions such as Topic Maps, whereby similar patents, classes, drawings, additional applications from the same inventors etc. are visualized.</li> <li>- through the selection of a search occasion. Depending on the occasion, only the relevant parameters are suggested or at least the more relevant parameters are suggested first (e.g., claims are very important for infringement search).</li> </ul>
<b>R12</b>	<p>The system must display the degree of similarity between the displayed patents to the users.</p> <p><b>Description:</b> It must become apparent to the user how big the similarity between the displayed patents is, for example in terms of a percentage. Additionally, there must be a hint explaining the calculation of the similarity. The similarity between a selected patent and other recommended patents can for example be displayed in the topic maps (cf. R11).</p>
<b>R13</b>	<p>The system must offer users the possibility to mark patents.</p> <p><b>Description:</b> It must be possible for users to mark patents</p> <ul style="list-style-type: none"> <li>- in order to give feedback for the system-generated relevance. This can be implemented in form of a mark „is interesting to me“/“is not interesting to me“.</li> <li>- in order to save them for later use (cf. R14).</li> </ul>
<b>R14</b>	<p>The system must allow users to save selected patents in lists.</p> <p><b>Description:</b> To be able to go back to interesting patents, it should be possible to save selected patents in lists. It should be possible to name the lists differently to structure the result sets.</p>
<b>R15</b>	<p>The system must offer users the possibility to save activities such as search formulations, added comments etc.</p>

	<b>Description:</b> It should be possible for users to save their searches in an archive. On doing so, added comments (and their authors), translations, links to patents should be saved and, depending on the sharing status, be available to other project members/employees of the company.
--	---

## 2b) Tooltip text/figures

<b>ID</b>	<b>Requirement</b>
<b>R16</b>	<p>The system must offer the users the possibility to mouse over a reference sign in the text and have the corresponding label or the reference sign list displayed.</p> <p><b>Description:</b> In order to quickly create a connection between the reference sign in the text and the corresponding component, the label (or the reference sign list) should be displayed as a tool tip on mouse over.</p>
<b>R17</b>	<p>The system must offer the users the possibility to mouse over a reference sign in the drawings and display the corresponding identifier.</p> <p><b>Description:</b> In order to quickly create a connection between the reference sign in the drawing and the corresponding label for the component, the corresponding label should be displayed in form of a tool tip on mouse over.</p>

### General requirement:

- R18: The system should present the information to the user in such a form that not much scrolling is necessary. Information should always be present – perhaps with a representer over which the information will be reachable quickly.
- The different goals of the search should be supported by the system.

### Nice to have:

- The system must offer the users the possibility to also search in non-patent literature.
- The system must offer users the possibility to use other research tools.
- The system must offer users the possibility to directly use figures from internet sources (e.g., a company's website) as input for the image search without having to download them from the website and re-upload them in the system.