

Fundamentals for automating due diligence processes in property transactions

Due diligence
in property
transactions

Philipp Maximilian Müller

Department of Civil Engineering,

*Unit of Real Estate Studies, Technische Universität Kaiserslautern,
Kaiserslautern, Germany*

Philipp Päuser

Architrave GmbH, Berlin, Germany, and

Björn-Martin Kurzrock

Department of Civil Engineering,

*Unit of Real Estate Studies, Technische Universität Kaiserslautern,
Kaiserslautern, Germany*

Received 30 September 2019

Revised 21 January 2020

3 March 2020

Accepted 6 March 2020

Abstract

Purpose – This research provides fundamentals for generating (partially) automated standardized due diligence reports. Based on original digital building documents from (institutional) investors, the potential for automated information extraction through machine learning algorithms is demonstrated. Preferred sources for key information of technical due diligence reports are presented. The paper concludes with challenges towards an automated information extraction in due diligence processes.

Design/methodology/approach – The comprehensive building documentation including $n = 8,339$ digital documents of 14 properties and 21 technical due diligence reports serve as a basis for identifying key information. To structure documents for due diligence, 410 document classes are derived and documents principally checked for machine readability. General rules are developed for prioritized document classes according to relevance and machine readability of documents.

Findings – The analysis reveals that a substantial part of all relevant digital building documents is poorly suited for automated information extraction. The availability and content of documents vary greatly from owner to owner and between document classes. The prioritization of document classes according to machine readability reveals potentials for using artificial intelligence in due diligence processes.

Practical implications – The paper includes recommendations for improving the machine readability of documents and indicates the potential for (partially) automating due diligence processes. Therefore, document classes are derived, reviewed and prioritized. Transaction risks can be countered by an automated check for completeness of relevant documents.

Originality/value – This paper is the first published (empirical) research to specifically assess the automated digital processing of due diligence reports. The findings are helpful for improving due diligence processes and, more generally, promoting the use of machine learning in the property sector.

Keywords Artificial intelligence, Machine learning, Property transactions, Due diligence, Digital building documentation, Document classes

Paper type Research paper

1. Research background and previous research

Information is key for making right decisions and achieving corporate goals (Bhatt, 2001, p. 70). Property is changing from an asset-intensive sector to an information-intensive sector. The majority of property data are collected in daily operations, e.g. regarding maintenance, letting, services charges or energy consumption. However, there is a fundamental lack of instruments for leveraging internal data and a risk of overlooking the essentials (Bawden and Robinson, 2009, pp. 182-184).

Traditionally, many property investors ground their decisions on a combination of intuition and retrospective data (Asaftei *et al.*, 2018). Although property investments are



generally assumed to follow a rational process many cognitive limitations may and do apply (Salzman and Zwinkels, 2017, pp. 81-83). Reports may be subject to overconfidence or confirmation bias, especially if the underlying data are difficult to access (Öhman *et al.*, 2013). Using available digital documents in the best possible way may help to obtain an unbiased view of a property investment. A global commercial property transaction volume of more than US\$800 billion in 2019 (Jones Lang LaSalle, 2020) underpins the relevance of receiving building information at the right time, in the right place, in the right format, in the right level of detail and above all, the right information (RICS, Royal Institution of Chartered Surveyors, 2017). It is still particularly difficult to assess the physical structure of a building due to shortcomings in data structure and quality (Wouda and Opdenakker, 2019, pp. 573-576). Different interests, systems and data requirements lead to regular building documentation losses over time (Bodenbender *et al.*, 2019, p. 177).

Due to the growing amount of data and the use of new technologies, it is now possible to make investment decisions more rational. For transactions, documents are usually made available in data rooms that ensure availability, timeliness and consistency. Conventional data rooms are used just for the purpose of a transaction. To use information further for building operation, there is an ongoing development from pure transaction data rooms to permanent data rooms. The aim is to provide information and documents without media breaks as well as to identify information unambiguously over all life cycle phases of a property (Bodenbender *et al.*, 2019).

Due diligence reports in property transactions usually relate to tax, technical, environmental, legal and commercial issues. Technical due diligence especially focuses on the assessment and evaluation of the structural condition and maintenance requirements (RICS, Royal Institution of Chartered Surveyors, 2009). Since construction and maintenance are associated with substantial costs, identified risks or doubts may trigger purchase price discounts, guarantees or even termination of negotiations. Generally, in just a few weeks a wide range of documents must be elaborately prepared and made available. Gaps in the documentation can only be identified with great effort. The quality and completeness of digital building documentation is increasingly becoming a factor as “deal maker” and “deal breaker” in transactions. One of the greatest potentials of artificial intelligence (AI) in property relates to services that support reporting (EY Real Estate, Vonovia, 2019).

Artificial intelligence can well assist in the classification of documents as key for the structuring of a building documentation. This requires a distinct set of document classes. Classification refers to the assignment of a document to a (predefined) class. Ideally, every document is assigned to one single document class. Previous achievements in automated classification already help to recognize, classify, name and sort digital documents in digital data rooms. Algorithms used are mostly based on machine learning (ML) and natural language processing (NLP) techniques. With different classification algorithms (Naïve Bayes, Support Vector Machine, Deep Learning) the error rate in recognizing and interpreting text in documents could continually be reduced. With methods of supervised learning, hit rates above 80% are already achieved (Bodenbender *et al.*, 2019; Russell and Norvig, 2016). The next step is information extraction. The goal is to recognize and extract predefined values and information such as entities or relationships (Jurafsky and James, 2019). In the long run, artificial intelligence could extract relevant information from documents to assist investment decision making.

Standards play a critical role in document classification, data exchange and information extraction. They ensure interoperability/compatibility, quality, optimization/variety reduction and information/measurement for evaluating processes (Hatto, 2013). From the point of view of research and development, standards are one of the most important instruments to further develop existing services with new technologies (such as artificial intelligence).

Certain approaches in the field of document and data standardization as well as data exchange in the property sector already exist.

OSCRE International, short for Open Standards Consortium for Real Estate, is a nonprofit, member-based organization which aims to develop, standardize and implement property data-exchange formats in commercial real estate. OSCRE has developed and published a series of XML data schemas for corporate real estate and investment applications (www.oscre.org).

The gif, Society of Property Researchers, Germany (Gesellschaft für Immobilien wirtschaftliche Forschung, www.gif-ev.de) is a nonprofit, member-based platform of the German property sector aiming to promote research and education as well as advancing standards to increase market transparency. The gif-IDA guideline for the exchange of property data (gif, 2017) aims to standardize the exchange of data between various parties. In March 2019, gif published “Standards for Configuring a Property Data Room and Document Management System 2.0” (gif, 2019). In addition to app. 100 document classes and a document index, it includes a mapping of the structure of the French partner association FIDJI.

FIDJI offers a free and open exchange of data format (Format d’Inter-échanges de Données Juridiques et Immobilières, or Financial and Property Data Interchange Format, www.format-fidji.org). FIDJI works on aligning their exchange of data format with gif and OSCRE. Currently, FIDJI counts 99 document classes (FIDJI, 2015).

Several organizations and research associations pursue similar goals in other countries, although they have not published sets of document classes yet:

- (1) UK: Royal Institution of Chartered Surveyors (RICS, www.rics.org), Society of Property Researchers (SPR, www.sprweb.com), Investment Property Forum (IPF, www.ipf.org.uk)
- (2) Netherlands: Vereniging Onroerend Goed Onderzoekers Nederland (VOGON, www.vogon.nl), Real Estate Data Exchange Association (REDEX, www.redex.nl)

The need for international harmonization of standards is recognized by many established organizations in the property sector, in particular with regard to floor measurement, performance measurement, accounting, reporting and valuation, such as BOMA (known for IPMS), MSCI, INREV, IFRS, IVSC, EPRA, NCREIF, PREA and RESO (for multiple listing services) [1].

Most of these initiatives aim to establish standards for structured and process-related data exchange between the parties in the life cycle of buildings. Automated digital processing of documents is seen as a subsequent field for research and development.

2. Aim of the research and research approach

The main added value of artificial intelligence will result from the targeted use of data. This research provides fundamentals for generating (partially) automated standardized technical due diligence reports by analyzing documents and extracting relevant information from building-related documents. Therefore, the paper seeks to identify unambiguous document classes and check the potential for digital processing of documents to promote the use of machine learning in the property sector.

The primary research question is:

What are the key requirements for information, documents and algorithms for an automated technical due diligence?

This is subject to the following underlying questions:

- (1) Which information is key for technical due diligence?

- (2) Which documents can the key information be sourced from?
- (3) Which challenges need to be addressed towards automated information extraction?

The research is grounded on 8,339 digital documents of 14 properties from eight (institutional) investors. The focus is on the structuring of documents as well as the identification of relevant documents for due diligence to support transaction processes. The documents shall be organized in a manageable number of unique and uniform document classes with a minimum of overlaps. The document classes are derived from the primary data including original due diligence reports and checked against industry standards such as the data room indices of [gif \(2019\)](#), [Drooms \(undated\)](#), [Fischer and Teufelsdorfer \(2015\)](#) and eight individual standards of the property owners in the data set. The document classes are then prioritized according to their relevance and availability in data rooms during property transactions. Further, documents across the document classes are principally checked for machine readability. Digital processing requires optical character recognition (OCR), which converts document scans, PDFs and image files into editable and machine-readable formats to identify characters and their position on a document page. Eventually, this will enable (partially) automated information extraction with machine learning.

3. Data

The data are provided by the PropTech Architrave [\[2\]](#). The primary data set including 8,339 digital documents of 14 properties from eight different owners is analyzed. The properties are located in nine German metropolitan regions and are predominantly used as offices. The building documentation of the properties includes building-related documents and company-related documents (see document classes in [Appendix](#)). An overview of the data set is presented in [Table 1](#). The table shows the amount of frequently occurring text-based files (e.g. .pdf, .xls(x), .doc(x)) as well as image-based files (e.g. .jpg and .tif(f)) and AutoCAD Drawing Database files (DWG) per property. Rare file formats such as .csv., .zip., .ppt, and .txt are subsumed in the column “other.”

The data set contains 7,330 pdf files (around 88% of all files) which are essential for classification and subsequent information extraction. Of the 8,339 documents, the majority typically relate to tenant and lease contract information (13.5%), plans (6.4%) as well as

Table 1.
Overview of building
documentation for each
asset based on type of
documentation (own
illustration, 2020)

No.	Location	Leasable area [sqm]	Text- based	Image- based	CAD- file	.xls/ .xlsx	.doc/ .docx	Other	No. of documents
1	Cologne	112.500	656	3	–	6	–	–	665
2	Cologne	2.800	1,381	21	21	15	1	1	1,440
3	Rhine Main area	13.800	473	2	–	5	–	–	480
4	Rhine Main area	35.000	1,270	466	17	37	1	1	1,792
5	Berlin	9.100	276	6	–	7	–	–	289
6	Berlin	16.200	362	–	–	6	–	–	368
7	Berlin	5.300	398	–	9	29	–	–	436
8	Magdeburg	7.500	439	62	–	1	–	–	502
9	Düsseldorf	7.500	222	–	–	–	4	–	226
10	Frankfurt	5.000	207	–	26	1	2	2	238
11	Munich	12.800	551	1	29	9	6	8	604
12	Munich	18.700	385	32	33	7	4	3	464
13	Ruhr area	5.000	359	–	13	38	13	2	425
14	Bonn	5.000	351	5	–	31	16	7	410
in total			7,330	598	148	192	47	24	8,339

building applications and building permits (5.2%). Image-based files including DWG files have a share of 8.9% of the total dataset. Image recognition, i.e. the identification of objects, buildings and patterns in photographs, is not part of this research. However, there are some promising approaches for the automated extraction of condition-specific visual characteristics from buildings (Koch *et al.*, 2018; Despotovic *et al.*, 2019). The majority of documents in the data set originate from the operating phase such as maintenance protocols, service and lease contracts or invoices for construction measures. The amount and quality of the documents and files is very heterogeneous depending on the owner and authors. The building sizes range from 2,800 sqm (office) to 112,500 sqm (warehouse). There is no direct correlation between the size of the asset (leasable area) and the number of documents. This can be explained by insufficient management, poor documentation or single/multi-tenant properties.

The structuring of the document classes also differs by organization. As mentioned, there is no industry-wide standard so far according to which property-related data and documents should be captured, managed, analyzed and exchanged. The initiatives by gif (gif, 2019), FIDJI and OSCRE are designed to become an industry-wide standard with a minimal number of document classes.

The research is further grounded on $n = 21$ technical due diligence reports of larger properties, the majority of which are owned by institutional investment companies. The original reports were checked for completeness and content according to guidance notes (RICS, 2011; Just and Stapenhorst, 2018; gif, 2019; Drooms, undated; Fischer and Teufelsdorfer, 2015). 85% of the technical due diligence reports are compliant with RICS (2011). Due to incomplete building documentation in data rooms and the homogeneity of properties, deviations from guidelines in particular cases are justified.

4. Results

4.1 Which information is key for technical due diligence?

Technical due diligence is carried out for the purpose of a technical assessment in property transactions. Contractors are usually sellers or potential buyers who request key information regarding the structural condition and maintenance requirements in the form of a report.

About 60–70% (app. 5,000–6,000) of the 8,339 documents in the data set are considered relevant for technical due diligence. The main common information sources of the technical due diligence reports in compliance with RICS (2011) are documents such as land register, public easement, construction and planning law or statutory compliance review. The analyzed due diligence reports have an average page count of around 30 pages, ranging from four pages to around 100 pages for a comprehensive technical due diligence report on a large property. The analysis shows that about 15–25 important key information items per page exist on technical due diligence contents such as general property description, structure, building physics, building fabric, external areas, building services, environmental and sustainability issues, capital expenditures (Capex) and operational expenditures (Opex). While some commonalities are found, components of a standardized technical due diligence reporting structure are beyond the scope of this paper. In total, about 500–600 key information items are included in an average technical due diligence report. For a partial automation of due diligence processes, key information items are presented. As an example, 16 key information items for technical due diligence are identified and mapped to the respective document class in Table 2. The columns to the right show up to three document classes (document A, B and C) which the information item can potentially be sourced from. Text-based documents are preferred at this point. The prioritized document classes in the Appendix feature machine readability scores. The key information item “year of construction” is most likely to be obtained from Doc A: 0072 planning application (see document classes in Appendix). If the “year of construction” is not contained in document A or the document is not available in the data room, document B and document C are further

Table 2.
List of key information
items and
corresponding
document classes to
source from (own
illustration, 2020)

No.	Key information	Unit/data field	Document A	Document B	Document C
1	address	text	0002	0003	0009
2	plot size	m ²	0072	0039	0080
3	land-use type	term	0058	0059	0060
4	year of construction	date	0072	0002	0080
5	year of renovation	date	0077	0139	0007
6	main use	term	0003	0007	0072
7	number of main floors	amount	0002	0072	0080
8	total gross floor area	amount	0002	0097	0098
9	total net lettable area	m ²	0200	0201	0080
10	heat supply	term	0280	0160	0249
11	contaminated sites	m ²	0111	0028	0029
12	parking spaces	amount	0002	0078	0083
13	age of elevator	date	0165	0184	0264
14	year of fire alarm system	date	0123	0255	0177
15	public land charges	text	0039	0042	0040
16	water consumption	[m ³ /year]	0283	0298	0235

alternatives containing the information. This assumes that, in practice, building documentation is often incomplete. This is a strong appeal to property companies to maintain and hand over complete building documentations across all life cycle phases.

The aim is to extract relevant information from the necessary documents in a fast and reliable way. The information extraction becomes more accurate when the documents are known in which the key information can best be found in terms of accessibility and machine readability.

4.2 Which documents can the key information be sourced from?

Due to the increasing migration of building documentation between life cycle phases, various software systems and document indices, there is a need for a holistic approach to structuring documents. Within the framework of this paper, 410 unique and uniform document classes (level 3) were derived from the data set (see [Appendix](#)). The analysis of 8,339 digital documents first comprises a systematic capture and structuring according to building and company-related documents (level 1). In addition, the building-related documents were classified regarding transaction, ownership, law, technical and organizational content, and the company-related documents were classified regarding management finance, reporting, company and commercial issues (level 2). The document classes thus derived are categorized as follows in [Table 3](#). The number of documents is indicated by n=x documents. In Table 3, “y” (yes) relates to text-based documents, “n” (no) to image-based documents and “yn” (yes and no) to documents that contain both text-based and image-based elements.

Basically, the building-related documents are of higher importance for technical due diligence as they include all technical data pertaining to the property. In ambiguous cases (e.g. third-party liability insurance), documents are declared as building-related. Documents relating to building technology and property and facility management are essential.

The document classes derived from the data were then compared with industry standards such as the data room indices of [gif \(2019\)](#), [Drooms \(undated\)](#), [Fischer and Teufelsdorfer \(2015\)](#) and eight individual standards of the owners in the data set. Supplementary documents of the existing standards were integrated into the document list in this research and redundancies were eliminated. The overlap-free document classes resulting from this approach enable a clear classification of documents, migration between standards and the structuring of documents for different use cases. The analysis shows that with an increasing number of building documentation, the number of document classes increases as well.

		Due diligence in property transactions
Building-related documents: <i>n</i> = 216 documents	Company-related documents: <i>n</i> = 194 documents	
- Property overview: <i>n</i> = 8 (y: 37.5%, <i>n</i> : 50.0%, yn: 13.5%)	- Company: <i>n</i> = 8 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Purchase: <i>n</i> = 13 (y: 100%, <i>n</i> : 0%, yn: 0%)	- Special purpose vehicle: <i>n</i> = 6 (y: 83.3%, <i>n</i> : 0%, yn: 16.7%)	
- Land: <i>n</i> = 16 (y: 62.5%, <i>n</i> : 12.5%, yn: 25.0%)	- Constitution: <i>n</i> = 2 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Tenure: <i>n</i> = 10 (y: 100%, <i>n</i> : 0%, yn: 0%)	- Shareholders' resolutions: <i>n</i> = 4 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Part ownership: <i>n</i> = 10 (y: 90%, <i>n</i> : 10%, yn: 0%)	- Management: <i>n</i> = 3 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Planning law: <i>n</i> = 11 (y: 72.7%, <i>n</i> : 0, yn: 27.3%)	- co.-related contracts: <i>n</i> = 4 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Building law: <i>n</i> = 11 (y: 63.6%, <i>n</i> : 0%, yn: 36.4%)	- Finances: <i>n</i> = 7 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Building information: <i>n</i> = 19 (y: 5.3%, <i>n</i> : 79.0%, yn: 15.7%)	- Joint ventures: <i>n</i> = 6 (y: 66.7%, <i>n</i> : 0%, yn: 33.3%)	
- Expert reports: <i>n</i> = 20 (y: 80%, <i>n</i> : 0%, yn: 20%)	- Financing: <i>n</i> = 17 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Fire protection: <i>n</i> = 11 (y: 54.5%, <i>n</i> : 18.2%, yn: 27.3%)	- Funds: <i>n</i> = 9 (y: 100%, <i>n</i> : 0%, yn: 0%)	
- Constr. documentation: <i>n</i> = 19 (y: 100%, <i>n</i> : 0%, yn: 0%)		Table 3. Overview of document class categories with number of documents (own illustration)
- Building tech. services: <i>n</i> = 51 (y: 98.0%, <i>n</i> : 0%, yn: 2.0%)		
- Lease contract: <i>n</i> = 37 (y: 94.6%, <i>n</i> : 0%, yn: 5.4%)		
- Property and facility mgmt.: <i>n</i> = 42 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Utilities: <i>n</i> = 6 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Insurance: <i>n</i> = 10 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Taxes and public charges: <i>n</i> = 16 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Other document: <i>n</i> = 4 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Market and environment: <i>n</i> = 10 (y: 20%, <i>n</i> : 10%, yn: 70%)		
- Controlling: <i>n</i> = 5 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Reporting: <i>n</i> = 6 (y: 100%, <i>n</i> : 0%, yn: 0%)		
- Exit: <i>n</i> = 9 (y: 88.9%, <i>n</i> : 0%, yn: 11.1%)		

The document classes in the [Appendix](#) may provide a basis for a classification standard for structuring property-related documents. The document classes do not overlap in terms of content to enable unambiguous classification. They facilitate information extraction by allowing algorithms to be focused on one class at a time. The classes each may require different algorithms since they contain different types of information. The document classes are continuously numbered (see [Appendix](#)). As they develop into a generally agreed standard, the numbering can be replaced by an appropriate form of indexing.

Documents need to be constantly categorized in specific structures as owners, applications and processes change. There are various company-specific indices, data room indices, document classes for operation and use as well as indices from literature. The structuring effort usually takes a lot of time, effort and cost. The document classes make it possible to structure documents largely without overlaps. At the same time, it is still a manageable number of document classes [3]. In property transactions, documents are usually classified into individual standards in data rooms. Non-overlapping classes enable the fast transfer to another standard, e.g. in the course of an ownership change. For all changes to the existing document classes, it must be ensured that a mapping to other standards is possible in order to achieve a successful migration.

To assess the completeness and validity of building documentation, general rules are developed for prioritized document classes according to relevance and machine readability of documents for technical due diligence as follows:

- (1) priority 1 (high) = includes at least one key information item for technical due diligence that is best available from this document in terms of machine readability (featuring the highest available machine readability score).
- (2) priority 2 (medium) = includes at least one key information item for technical due diligence that is machine-readable.

- (3) priority 3 (low) = may not be machine-readable but still includes at least one key information item for technical due diligence.
- (4) priority 4 (–) = may or may not include a key information item for technical due diligence but is inferior in terms of machine readability to at least one other document class offering the same information item.

The analysis is based on an industry-standard optical character recognition (OCR) software. Prior to the OCR process, the structure of the document is analyzed by the software and areas containing text, images and tables are recognized. OCR technology is then used to convert document images into editable and machine-readable text. It shows that the documents in this domain are rather difficult to read by machine. For instance, a large proportion (>80%) of maintenance protocols is marked with handwritten notes. Several findings apply to all types of documents:

- (1) Document condition: Paper documents which have been exposed to frequent use are much less recognizable by machine than fresh printouts. Handwritten notes further reduce machine readability substantially.
- (2) Scan quality: Color depth, resolution and density range influence the quality of a scan and thus machine readability; the same counts for documents with “noise” (i.e. scattered black dots), blurred or uneven letters or lines and shifted tables.
- (3) Document structure: The less structured a document, the more difficult is a meaningful extraction or the more documents are needed to train extraction algorithms.

The analysis shows that these characteristics limit the potential of digitization for machine readability. The quality of the documents has a tremendous impact on the optical character recognition quality. Some details or handwritten notes could hardly be identified even by visual examination. In particular, when handwritten notes interfere with typed characters, optical character recognition frequently fails. Digitally created documents and contracts, such as service contracts or utility contracts, are generally easier to read by machine.

The list in the [Appendix](#) prioritizes the document classes by the share of documents that are generally easy to read by machine. The machine readability scores are specified as follows:

- (1) score 1 (high) = digitally generated document with character recognition.
- (2) score 2 (medium) = digitized documents with few images and minor handwritten notes.
- (3) score 3 (low) = digitized documents with many images and handwritten notes as well as multiply scanned documents of low quality.

An example of an optical character recognition analysis is shown in [Figure 1](#): The document, a maintenance report for an elevator (document class 0184, [Appendix](#)) was checked for machine readability. The areas marked in red [logo, stamp, signature] are not recognized and therefore difficult or impossible to read by machine. The two blue marked areas [table] in the middle of the document are recognized as tables. If the font is easy to read, e.g. featuring a high contrast, the recognized area can be exported in table format and further processed digitally. The green text areas [text] indicate good machine readability and offer the best potential for information extraction. The document class (0184) contains important information regarding elevator maintenance (priority 1). However, it is not fully machine-readable (machine readability score: 2). There are minimal errors in the interpretation of

Zugelassene Überwachungsstelle
Aufzüge

-315 -189022 -20720 -421 .doc

Straße 1A, 82152 Planegg b. München

GmbH
Industrie, Bau und Immobilien
Außenstelle München
Straße 1A
82152 Planegg b. München
Telefon +49.89 -0
Telefax +49.89 -47

Kontakt
Tel. direkt
E-Mail

Seite 1 von 1

logo

text

Prüfbescheinigung

Zwischenprüfung

Prüfgrundlage: gem. § 15 BetrSichV

Objektstandort: 80807 München,

Objektart / Anlage: Personenaufzug

Fabrik-Nr.: 207

Betreiber:

Verwendete Messgeräte:

Technische Angaben

Baujahr: 2005

Wartungsfirma:

Errichtungsgrundlage: EN 81-1

Antrieb / Aufhängung: Treibseil / 2:1

Hersteller: OSMA-Aufzüge Albert Schenk GmbH & Co. KG

Tragfähigkeit: 1080 kg / 14 Pers

Haltestellen / Zugangsstellen: 4 / 4

Geschwindigkeit: 1,00 m/s

Förderhöhe: 11,56 m

Ergebnis der Prüfung: Geringe Mängel

Id.-Nr.	Mangelbeschreibung	Mangelnummer	Bewertung
1	Fahrkorb Beide Spiegel weisen Beschädigungen auf - Verletzungsgefahr	319	geringfügig
2	Ordnungsprüfung/Prüfung der technischen Unterlagen Sicherheitstechnische Bewertung/Gefährdungsbeurteilung (zur Ermittlung der Prüf Fristen) konnte während der Prüfung nicht eingesehen werden	804	geringfügig

Mängelbeseitigung bis: Nachprüfung nicht erforderlich

Bemerkungen:
Bezogen auf das Gefahrenfeld Absturz bestehen gegen den Weiterbetrieb keine sicherheitstechnischen Bedenken. Die Mängel gem. Auflistung der Prüfbescheinigung sind zu beheben.

signature

stamp

Nächste Prüfung: 01.2015

Datum der Prüfung:

122

315

189022

20720

421

www.com

Sitz Stuttgart, Amtsgericht Stuttgart

Bankverbindung
Commerzbank AG
Kto.-Nr. 9 82 600 800 00
Landesbank Baden-Württemberg
Kto.-Nr. 2 82 600 501 01

Vorsitzender des Aufsichtsrates
Geschäftsführer (Vorsitzender)

Figure 1.
Machine readability
analysis with optical
character recognition
(own illustration, 2020)

letters and numbers which can be processed manually afterwards. In the course of automation, such minimal errors must be corrected. It shows that validity, timeliness, objectivity and stability are essential quality requirements for documents which help to increase the performance of digital processing.

Overall, 410 document classes in 32 categories were analyzed (Table 3). The analysis reveals that a substantial part of relevant digital building documents is insufficient for automated information extraction (compare list in Appendix). The availability and content of documents vary greatly from owner to owner, from document to document and between document classes. Standardized documents are easier to be read by machines, such as:

- (1) concession documents,
- (2) official certificates,
- (3) infrastructure provision,
- (4) building permit documents,
- (5) commercial register extract,
- (6) land register extract.

App. 250 (60–70%) of all document classes are relevant for technical due diligence in different ways. The exclusion of the remaining 160 classes is due to:

- (1) lack of content pertaining to technical due diligence,
- (2) lack of machine readability (plan, photos, map),
- (3) inferior machine readability of documents against other document classes.

The distribution of document classes according to priorities for technical due diligence is shown in Table 4. 20.7% of document classes have high priority (80–90 classes), 23.2% have medium priority (90–100 classes), 23.2% have low priority (60–70 classes) and 40.2% have zero priority (160–170 classes). Documents in which the essential key information items are predominantly in text form (e.g. 0039: land register extract) are declared as text-based documents with “y” (yes). Image-based documents (e.g. 0082, floor layout plan) are labeled with “n” (no text-based classification possible). Documents that contain both text-based and image-based elements (e.g. 0022, official site plan) are marked with “yn” (yes and no). As intended, the machine readability scores do significantly correlate with the proportion of text-based documents. The machine readability of text-based documents is generally promising due to the use of standardized electronic fonts, a multitude of pre-trained algorithms and data sets as well as software systems that frequently rely on dictionary support. The score improves with the share of text-based documents. On average, the machine readability score is 1.82 for the document classes with priority 1, 2 and 3 and is therefore predominantly insufficient for digital processing. The more text-based documents in a document class, the better the machine readability. Overall, the proportion of text-based documents is more than sufficient (85.4%). However, there are documents like protocols, which are text-based but much more difficult to process digitally due to frequent handwritten notes. In priority 2 (medium), there are several as-built plans (Appendix: document classes 0081-0095) where

Table 4.
Distribution of document classes by priorities and corresponding machine readability score (own illustration, 2020)

Priority	Document classes		Machine readability score		
	Number	Text-based [%]	Average	Median	Modal value
1 (high)	80–90 DC	84.3%	1.79	2.00	2.00
2 (medium)	90–100 DC	76.7%	1.89	2.00	2.00
3 (low)	60–70 DC	83.6%	1.75	2.00	2.00
4 (none = 0)	160–170 DC	91.6%	–	–	–
in total	410 DC	85.4 %	1.82	2.00	2.00

algorithms for optical character recognition are not applicable. Image recognition is required to extract information from these documents.

It is suggested that the document classes can be used to compare classified documents with a target list. Such a comparison enables automated and reliable identification of gaps or redundancies in the documentation, which is an important basis for structured due diligence processes. This information is immediately available and missing documents can be obtained at an early stage in property transactions.

Which challenges need to be addressed towards automated information extraction?

The goal of information extraction is to find, link and extract relevant information from text ignoring irrelevant information. Digital building documents generated in the property life cycle serve as input. The output would be an instantiated template filled with the respective text fragments in the form of an automated technical due diligence report.

In managing property, the capture, provision and updating of building data is a costly and important task. Understanding data management is critical for reducing user error and improving efficiency. Without understanding how data are captured, stored, updated or used, organizations face significant risk in their day-to-day operations. However, document and information losses regularly occur in the property life cycle due to ownership changes, different systems and parties involved. Especially in the course of transactions, it is all the more important to carry out a structured data transfer. The data transfer from the previous to a new owner is frequently subject to an inefficient and delayed process (Saul and Baum, 2019). Through digital processing, human capacities can be focused on value-adding processes and efficiency can be increased. This may change in future with a comprehensive digital building documentation.

In the course of this research, we find the following challenges that still need to be addressed towards digital processing and (partially) automated due diligence processes:

- (1) If PDF files are scanned images, these are hard to read for machines; owners must ensure that documents are digitized properly (contrast, alignment etc.) and optical character recognition is applied; remaining errors must be corrected (manually) to prevent erroneous data in processing systems,
- (2) automated systems should be able to highlight potential alpha (false positive: wrong information) and beta (false negative: missing information) errors; otherwise, reports will ultimately create a false sense of security,
- (3) documents of the same type have manifold layouts (fonts, font size, partitions, tables); this can generally be solved by machine learning algorithms, at least if text content prevails,
- (4) nonetheless, documents that are standardized to some extent will facilitate automated information extraction,
- (5) illustrations, signatures and handwritten notes frequently cause problems, especially when interfering with typed characters,
- (6) even important documentation is often missing: e.g. reconstruction/conversion, fire protection reports.

It is highly recommendable that data managers take ownership of data, similar to risk owners in risk management. This is particularly true for building-related data that will eventually be migrated into external systems (e.g. building information modeling). This may promote data quality in a wider sense and prevent erroneous use of data. Clearly, data management will be more and more supported by artificial intelligence. However, artificial intelligence can hardly substitute responsibility of human professionals. In fact, this may become a major business in

the future. When reports are generated (semi-)automatically, as expected in the future, the need to ensure data quality will become even more important.

5. Limitations

5.1 Requirements and scope

Obviously, documents in data rooms or document management systems must meet certain requirements regarding content and quality. To increase hit rates in document classification and to successfully use data through information extraction, more standardized documents will be essential. Therefore, documents should be enriched with meta information (e.g. xml attachment to PDF) when created or modified. These aspects are the subject of further research.

In the case of a vendor due diligence, all resources can usually be accessed in an adequate time frame for comprehensive inspection with extensive accessibility and documentation. If required, additions can be requested quickly. In an investor due diligence, limited resources are available within a limited period of time. It is expected that the building documentation as basis for digital processing may differ by use case (vendor/investor technical due diligence) to some extent. This aspect may be reflected in future research.

Different types of property and specifications (e.g. green buildings) may require other technical due diligence contents. New processes and applications in the construction and property sector may also call for additional document classes.

The document classes derived in this paper seek to represent a standard for as many building types as possible, which will have to be supplemented or modified in individual cases. The migration capability can eventually be achieved by introducing a meta-standard.

5.2 International standardization

The results are grounded on data from 14 properties of eight institutional investors. Even these showed quite heterogeneous approaches towards managing documents. It is necessary to promote standardization and digital processing more deeply across the industry. The works from gif ([gif, 2019](#)), FIDJI, OSCRE and RICS ([RICS, 2009, 2011, 2017](#)) are important steps in this direction taking an international perspective. This may add to the growing demand for transparency in international property markets ([Newell, 2016](#)).

6. Conclusions

The overall picture that emerges from this research is that there are promising approaches and advances for the digitalization of technical due diligence processes through artificial intelligence. Risks in transaction processes can be countered by an automated identification and verification of relevant documents at a reduced cost. As noted above, however, there is still some way to go for using artificial intelligence for digital processing in transaction processes.

The empirical results lead to the following practical recommendations:

- (1) A high level of document availability, completeness, uniformity, timeliness and machine readability is key to the success of automated information extraction. Since the analysis of the building documentation has shown that 46.1% of relevant due diligence documents are more or less unsuited for automated information extraction due to limited machine readability scores of 2 or 3, an increase in document quality and exchange of digital (rather than digitized) documents is vital for the digital transformation of the property sector. Documents and data are often used in isolation,

resulting in different data storage and file names between business units. Common documentation standards can be interpreted equally by multiple groups.

- (2) Another essential requirement for information extraction are suitable document classes. A set of document classes must neither be too small nor too bulky. Priorities of document classes for specific purposes, such as technical due diligence, can assist with data procurement. This allows to focus on documents that are most helpful with providing relevant information at the lowest cost. It might be useful to expand the existing document classes selectively, since document classes are rather dynamic than static. With all additions or changes to the existing document classes, it is important to ensure a consistent mapping to other standards. Steps towards an international standardization in document management are already being taken (gif, 2019, FIDJI, OSCRE and RICS, 2009, 2011, 2017). The standardization of building documentation and data exchange is necessary for harmonization and transparency on a global scale to ensure smooth transfer of documents and information in property transactions.
- (3) Prioritizing the machine readability of technical due diligence documents and digital processing may empower companies for the (partial) automation of due diligence processes. Due diligence processes will not be fully digitized until there are complete digital processes in property management and a standardized set of documents that can be read by machine. Digitally generated documents can provide necessary information with metadata. This will also improve building documentation.
- (4) For the (partial) automation of due diligence a standardized reporting structure will be essential. In literature there are some approaches for due diligence checklists. A standardized due diligence structure does not yet exist but could provide a number of benefits. In the long run with the use of artificial intelligence, a standardized, automatically generated report could help to ground investment decisions on documents analyzed. A standard report structure for technical due diligence, based on the key information identified in this paper, will be subject of further research.

The results offer various potentials for further research. For instance, the findings about machine readability of building documentation show that text-intensive areas such as legal and tax due diligence are generally suitable for (partial) automation. While the focus of this research is on document standardization and analysis of technical due diligence information, it is equally important to understand where the data originated. The automation of document generation (smart contracts) and the input process (digital interfaces) are further steps to increase data efficiency. Moreover, since due diligence providers increasingly operate internationally and reach a wider range of clients, different international perspectives could be reflected to improve (and eventually harmonize) national standards.

Notes

1. BOMA = Building Owners and Managers Association, INREV = European Association for Investors in Non-Listed Real Estate Vehicles, IFRS = International Financial Reporting Standards, IVSC: International Valuation Standards Council, EPRA = European Public Real Estate Association, NCREIF = National Council of Real Estate Investment Fiduciaries, PREA = Pension Real Estate Association, RESO = Real Estate Standards Organization.
2. Architrave provides an ML data room technology; the asset management platform comprises data on >4,200 assets of >90 bn. Euros in value; <https://www.architrave.de/en/>. Last checked 05 December 2019.

3. Ideally, every document should be assignable to just one single class. On the other hand, with a large number of classes users will lose oversight; even if machine learning algorithms manage to classify documents automatically. Investors and data room providers indicate that a set of 1,000 or more document classes would be too bulky and a significant portion of it would practically remain empty.

References

- Asaftei, G.M., Doshi, S. and Means, J. (2018), "Getting ahead of the market", *Urban Land Magazine by Urban Land Institute*, available at: <https://urbanland.uli.org/economy-markets-trends/getting-ahead-of-the-market> (accessed 05 December 2019).
- Bawden, D. and Robinson, L. (2009), "The dark side of information. Overload, anxiety and other paradoxes and pathologies", *Journal of Information Science*, Vol. 35 No. 2, pp. 180-191.
- Bhatt, G.D. (2001), "Knowledge management in organizations. Examining the interaction between technologies, techniques, and people", *Journal of Knowledge Management*, Vol. 5 No. 11, pp. 68-75.
- Bodenbender, M., Kurzrock, B.-M. and Müller, P.M. (2019), "Broad application of artificial intelligence for document classification, information extraction and predictive analytics in real estate", *Journal of General Management, Special Issue on Real Estate in the Digital Era*, Vol. 44 No. 3, pp. 170-179.
- Despotovic, M., Koch, D. and Leiber, S. (2019), "Automatic extraction of condition-specific visual characteristics from buildings", paper presented at *26th Annual European Real Estate Society Conference*, July 2019, Cergy-Pontoise, France, available at: <https://eres.architecture.net/doc/oai-eres-id-eres2019-284> (accessed 16 October 2019).
- Drooms (Ed.) (undated), *Drooms Due-Diligence-Index™*, Frankfurt, available at: <https://drooms.com/de/due-diligence-checkliste?whitepaper=1671> (accessed 06 December 2019).
- EY Real Estate, Vonovia (Ed.) (2019), "Asset management study 2019", available at: [https://www.ey.com/Publication/vwLUAssets/ey-real-estate-asset-management-studie-2019/\\$FILE/ey-real-estate-asset-management-studie-2019.pdf](https://www.ey.com/Publication/vwLUAssets/ey-real-estate-asset-management-studie-2019/$FILE/ey-real-estate-asset-management-studie-2019.pdf) (accessed 10 October 2019).
- FIDJI, Format d'Inter-échange de Données Juridiques et Immobilières (Ed.) (2015), "Plan de classement documentaire - FIDJI", available at: http://format-fidji.org/wp-content/uploads/2018/03/Plan_de_classement_Documentaire_FIDJI.pdf (accessed 06 December 2019).
- Fischer, P. and Teufelsdorfer, H. (2015), "Due-diligence-prozess", in Teufelsdorfer, H. (Ed.), *Handbuch Immobilientransaktionen*, Linde, Vienna.
- gif, Society of Property Researchers, Germany (Ed.) (2017), *Gif-IDA - Guideline of Real Estate Data Exchange*, gif, Gesellschaft für immobilienwirtschaftliche Forschung e.V., Wiesbaden.
- gif, Society of Property Researchers, Germany (Ed.) (2019), *Standards for Configuring a Property Data Room and Document Management System 2.0*, gif, Gesellschaft für immobilienwirtschaftliche Forschung e.V., Wiesbaden.
- Hatto, P. (2013), "Standards and standardisation", *European Commission – Directorate-General for Research and Innovation*, Brussels, available at: http://ec.europa.eu/research/industrial_technologies/pdf/handbook-standardisation_en.pdf (accessed 06 December 2019).
- Jones Lang LaSalle (JLL) Global Research (2020), "Global market perspective, February 2020", available at: www.jll.de/en/trends-and-insights/research/global-market-perspective-february-2020 (accessed 03 March 2020).
- Jurafsky, D. and James, H.M. (2019), "Information extraction", in Jurafsky, D. and James, H.M. (Eds.), *Speech and Language Processing*, Stanford, available at: <https://web.stanford.edu/~jurafsky/slp3/> (accessed 11 October 2019).
- Just, T. and Stapenhorst, H. (2018), *Real Estate Due Diligence – a Guideline for Practitioners*, Springer International Publishing, Cham.

-
- Koch, D., Despotovic, M., Sakeena, M., Döller, M. and Zeppelzauer, M. (2018), "Visual estimation of building condition with patch-level ConvNets", *Proceedings of the 2018 ACM Workshop on Multimedia for Real Estate Tech*, ACM, New York, NY, pp. 12-17.
- Newell, G. (2016), "The changing real estate market transparency in the European real estate markets", *Journal of Property Investment and Finance*, Vol. 34 No. 4, pp. 407-420.
- Öhman, P., Söderberg, B. and Westerdahl, S. (2013), "Property investor behaviour: qualitative analysis of a very large transaction", *Journal of Property Investment and Finance*, Vol. 31 No. 6, pp. 522-544.
- RICS, Royal Institution of Chartered Surveyors (Ed.) (2017), *Global Trends in Data Capture and Management in Real Estate and Construction*, RICS Insight Paper, London.
- RICS, Royal Institution of Chartered Surveyors Europe (Ed.) (2011), "RICS best practice & guidance note for technical due diligence of commercial, industrial & residential property in continental Europe", available at: <https://www.rics.org/globalassets/rics-website/media/upholding-professional-standards/sector-standards/building-surveying/technical-due-diligence-of-commercial-industrial-and-residential-property-in-continental-europe-1st-edition-rics.pdf> (accessed 02 October 2019).
- RICS, Royal Institution of Chartered Surveyors Oceania (Ed.) (2009), "Technical due diligence of commercial & industrial property Australia", available at: <https://sgaproperty.com/docs/RICS-Due-Diligence-Guidance-Note.pdf> (accessed 02 October 2019).
- Russell, S.J. and Norvig, P. (2016), *Artificial Intelligence: A Modern Approach*, Pearson, Boston.
- Salzman, D. and Zwinkels, R.C.J. (2017), "Behavioral real estate", *Journal of Real Estate Literature*, Vol. 25 No. 1, pp. 77-106.
- Saull, A. and Baum, A. (2019), *The Future of Real Estate Transactions*, Saïd Business School University of Oxford, Oxford, available at: https://www.sbs.ox.ac.uk/sites/default/files/2019-03/FoRET-ReportFull_1.pdf (accessed 11 October 2019).
- Wouda, H.P. and Opendakker, R. (2019), "Blockchain technology in commercial real estate transactions", *Journal of Property Investment and Finance*, Vol. 37 No. 6, pp. 570-579.

Appendix

List of Document Classes

The document classes are continuously numbered (class label). As they develop into a generally agreed standard, the numbering can be replaced by an appropriate form of indexing.

Text-based Classification/Subject of technical due diligence (TDD): y = yes, n = no, yn = yes and no (e.g. 0022, official site plan, which contains text and image-based elements)

Prioritization of document classes according to relevance and machine readability of documents for technical due diligence as follows: Priority 1 = high, 2 = medium, 3 = low, 4 = none

The machine readability scores are specified as follows: 1 = high, 2 = medium, 3 = low

Corresponding author

Philipp Maximilian Müller can be contacted at: philipp.m.mueller@bauing.uni-kl.de

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
<i>Building-related documents</i>					
0001 0002 0003 0004 0005 0006 0007 0008 0009 0010 0011 0012 0013 0014 0015 0016 0017 0018 0019 0020 0021	Property overview	<i>n</i>	<i>n</i>	4	-
	Data, document exchange folder	<i>y</i>	<i>y</i>	1	1
	Brochure	<i>y</i>	<i>y</i>	1	1
	Property description	<i>n</i>	<i>y</i>	2	3
	Property photographs	<i>n</i>	<i>n</i>	4	3
	Aerial photograph	<i>n</i>	<i>y</i>	3	3
	Visualization	<i>y</i>	<i>y</i>	1	1
	Type of use	<i>yn</i>	<i>y</i>	2	2
	Point of contact	<i>y</i>	<i>y</i>	1	1
	Purchase	<i>y</i>	<i>y</i>	1	1
	Exposé	<i>y</i>	<i>y</i>	2	2
	Purchase presentation	<i>y</i>	<i>n</i>	4	2
	Purchase price estimate	<i>y</i>	<i>n</i>	4	2
	Agent purchase	<i>y</i>	<i>n</i>	4	2
	Data room	<i>y</i>	<i>n</i>	4	3
0022 0023 0024 0025 0026 0027 0028 0029 0030 0031 0032	Q&A	<i>y</i>	<i>y</i>	3	1
	Sale and purchase agreement (SPA)	<i>y</i>	<i>y</i>	2	2
	Reference document	<i>y</i>	<i>y</i>	3	2
	Transfer of ownership	<i>y</i>	<i>n</i>	4	3
	Declaration of release	<i>y</i>	<i>n</i>	4	2
	Further purchase-related documents	<i>y</i>	<i>y</i>	2	2
	Correspondence attorney and notary	<i>y</i>	<i>n</i>	4	2
	Real estate transfer tax (RETT)	<i>y</i>	<i>n</i>	4	2
	Land	<i>yn</i>	<i>y</i>	2	3
	Official site plan	<i>n</i>	<i>y</i>	2	3
	Cadastral plan	<i>y</i>	<i>y</i>	2	1
	State-specific characteristic	<i>y</i>	<i>y</i>	2	2
	Land register	<i>y</i>	<i>y</i>	2	2
	Intervening space certification	<i>y</i>	<i>y</i>	1	2
	Superstructures	<i>yn</i>	<i>y</i>	1	2
	Register of contaminated sites	<i>y</i>	<i>y</i>	1	1
	Remediation of contamination	<i>y</i>	<i>y</i>	1	1
	Ordnance	<i>y</i>	<i>y</i>	1	1
	Foundation and geotechnical report	<i>y</i>	<i>y</i>	2	2
	Guideline land value map	<i>yn</i>	<i>n</i>	4	3
<i>(continued)</i>					

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0033	Existing trees	yn	y	3	2
0034	Preparation for building	y	y	3	1
0035	Disposal balance, demolition work	y	y	3	1
0036	Flood protection concept	n	y	2	3
0037	Investor contract, archaeological heritage management	y	y	2	2
	Tenure				
0038	Extract from the commercial register relating to the owner	y	n	4	2
0039	Land register extract	y	y	1	1
0040	Grant of entry in the land register	y	y	3	2
0041	Easements	y	y	2	2
0042	Public land charges	y	y	2	2
0043	License agreement	y	y	3	2
0044	Release of pledge	y	y	3	2
0045	Heritable building right contract, ground lease (UK)	y	n	4	2
0046	Ground rent, proof of payment	y	n	4	2
0047	Neighborhood agreement	y	y	3	2
	Part ownership				
0048	Part ownership	y	n	4	2
0049	Part ownership deed	y	n	4	2
0050	Community order	y	n	4	1
0051	Partition plan	n	n	4	3
0052	Condominium ownership deed	y	y	4	2
0053	Condominium owner's association records	y	n	3	2
0054	Business plan	y	n	4	2
0055	Condominium owner's association maintenance reserve	y	n	4	2
0056	Condominium owner's association property mgmt. contract	y	n	4	1
0057	House rules	y	n	4	1
	Planning law				
0058	Land use plan	yn	y	3	3
0059	Local development plan	yn	y	1	2
0060	Information relating to planning law	y	y	2	2
0061	Protection of ancient monuments	y	y	1	1
0062	Site servicing costs	y	y	3	1
0063	Correspondence on land improvement	y	y	3	1
0064	Planning agreement	y	y	2	1
0065	Land reallocation area, general improvement area	y	y	2	2

(continued)

Due diligence
in property
transactions

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0066	Preservation order	y	y	2	2
0067	Official certificates	y	y	3	1
0068	Rainwater drainage	yn	y	3	2
0069	Building law				
0069	Outline planning application	yn	y	3	2
0070	Tender documents	y	y	3	2
0071	Outline building permit	y	y	3	2
0072	Planning application	yn	y	1	2
0073	State building control	y	y	2	1
0074	Building permit documents	yn	y	2	2
0075	Demolition consent, demolition permit	y	y	3	2
0076	Official approval	y	y	2	2
0077	Use approval, change of use approval	y	y	1	1
0078	Parking space proof, redemption, compensation	y	y	1	2
0079	Structural engineer's report, approval, calculations	yn	y	2	3
0080	Building information				
0080	Building description	y	y	1	1
0081	As-built document, site plan, distance space	n	y	1	3
0082	Floor layout plan	n	y	1	3
0083	Parking space plan	n	y	2	3
0084	Reflected ceiling plan	n	y	3	3
0085	As-built elevations	n	y	2	3
0086	As-built sections	n	y	2	3
0087	Detailed plan	n	y	2	3
0088	As-built plan superstructure	n	y	2	3
0089	As-built plan heating	n	y	2	3
0090	As-built plan ventilation	n	y	2	3
0091	As-built plan air conditioning	n	y	2	3
0092	As-built plan sanitary installations	n	y	2	3
0093	As-built plan electrical	n	y	2	3
0094	As-built plan media	n	y	2	3
0095	As-built plan pipes and cables	n	y	2	3
0096	Locking system, keys, code cards	yn	n	4	2
0097	Schedule of areas	yn	y	1	2
0098	Measurement survey	yn	y	1	3

(continued)

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0069	Expert reports				
0100	Technical due diligence report	y	y	1	1
0101	Commercial due diligence report	y	y	1	1
0102	Legal due diligence report	y	n	3	1
0103	Tax due diligence report	y	n	4	1
0104	Financial due diligence report	y	n	4	1
0105	Environmental due diligence report	y	n	4	1
0106	Building survey, preservation of evidence	y	y	1	1
0107	Drinking water report	y	y	1	1
0108	Sound insulation report	y	y	1	1
0109	Underground services report	y	y	1	1
0110	Testing water tightness of drainage pipes	yn	y	1	1
0111	Contaminated sites report, pollutants	y	y	1	2
0112	Inspection reports building	yn	y	1	2
0113	Environmental certification	y	y	1	2
0114	Energy performance certificate (EPC)	yn	y	1	2
0115	Thermal insulation certificate	yn	y	1	2
0116	Cooling load determination	y	y	3	2
0117	Building certification	y	y	1	2
0118	Valuation, market value report	y	y	3	1
	Fire protection				
0119	Fire protection requirements	y	y	1	1
0120	Fire protection approval, fire prevention	y	y	1	2
0121	Fire protection regulation	y	y	1	2
0122	Fire protection concept	yn	y	1	3
0123	Fire protection survey	y	y	1	2
0124	Approval of deviation request	yn	y	1	2
0125	Test report sprinkler system	y	y	1	2
0126	Floor layout plans for fire brigade use	n	y	2	3
0127	Escape route plan	yn	y	2	3
0128	Interaction tests	y	y	3	2
0129	Fire control matrix	n	y	3	3
	Construction documentation				
0130	Technical planer contract	y	y	3	1
0131	General contractor agreement	y	n	4	1
(continued)					

Due diligence
in property
transactions

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0132	Maintenance plan	y	y	1	2
0133	Completed construction projects	y	y	1	2
0134	Ongoing construction projects	y	y	1	2
0135	Planned construction projects	y	y	1	2
0136	Deferred maintenance, defects	y	y	1	2
0137	Completed maintenance, confirmation of defect remediation	y	y	1	2
0138	Building invoices	y	y	2	2
0139	Official approval report	y	y	2	2
0140	Official approval report at completion	y	y	2	2
0141	TUV certificates	y	y	2	1
0142	Notification of completion	y	y	3	2
0143	Guarantee bond	y	y	2	2
0144	Guarantee overview	y	y	2	2
0145	Site managers declaration	y	y	3	2
0146	Specialized company declaration	y	y	3	2
0147	Building inspectorate approval	y	y	3	2
0148	Building products	y	y	2	2
0149	Building technical services	y	y	1	1
0150	Description of building services	yn	y	1	2
0151	Overview of building services	y	y	2	2
0152	Operation manual heating	y	y	2	2
0153	Operation manual ventilation	y	y	2	2
0154	Operation manual air conditioning	y	y	2	2
0155	Operation manual sanitary installations	y	y	2	2
0156	Operation manual fire protection system	y	y	2	2
0157	Operation manual elevators and escalators	y	y	2	2
0158	Operation manual electrical	y	y	2	2
0159	Operation manual UPS (uninterrupted power supply)	y	y	2	2
0160	Operation manual other	y	y	2	2
0161	Test report heating	y	y	1	2
0162	Test report ventilation	y	y	1	2
0163	Test report air conditioning	y	y	1	2
0164	Test report sanitary installations	y	y	1	2
0165	Test report fire protection system	y	y	1	2
0166	Test report elevator and escalators	y	y	1	2
	Test report electrical	y	y	1	2

(continued)

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0167	Test report UPS (uninterrupted power supply)	y	y	1	2
0168	Test report other	y	y	1	2
0169	Maintenance schedule	y	y	1	1
0170	Maintenance report heating	y	y	1	2
0171	Maintenance report ventilation	y	y	1	2
0172	Maintenance report air extraction system	y	y	1	2
0173	Maintenance report forced air cooling plant	y	y	1	2
0174	Maintenance protocol air conditioning	y	y	1	2
0175	Maintenance report sanitary	y	y	1	2
0176	Maintenance report fire protection technology	y	y	1	2
0177	Maintenance report fire alarm system	y	y	1	2
0178	Maintenance report fire damper	y	y	1	2
0179	Maintenance report fire doors	y	y	1	2
0180	Maintenance report fire protection gates	y	y	1	2
0181	Maintenance report hydrants	y	y	1	2
0182	Maintenance report fire report sprinkler system	y	y	1	2
0184	Maintenance report elevator and escalator	y	y	1	2
0185	Maintenance report electrical	y	y	1	2
0186	Maintenance report electroacoustic system	y	y	1	2
0187	Maintenance report safety lighting	y	y	1	2
0188	Maintenance report lightning protection	y	y	1	2
0189	Maintenance report UPS (uninterrupted power supply)	y	y	1	2
0190	Maintenance report emergency power generator	y	y	1	2
0191	Maintenance report building control system	y	y	1	2
0192	Maintenance report CO warning system	y	y	1	2
0193	Maintenance report ventilation and air-conditioning system	y	y	1	2
0194	Maintenance report smoke and heat exhaust ventilation system	y	y	1	2
0195	Maintenance report grease separator	y	y	1	2
0196	Maintenance report engines	y	y	1	2
0197	Maintenance report barriers	y	y	1	2
0198	Maintenance report drinking water test	y	y	1	2
0199	Maintenance report others	y	y	1	2
<i>Lease contract</i>					
	Central documents				
0200	Rent roll, tenancy schedule	y	y	3	2
0201	Occupancy plan, plan of lettable data sheet	y	n	4	-
<i>(continued)</i>					

Due diligence
in property
transactions

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0203	Sample lease contract	y	n	4	-
0204	Description of tenant fit-out	y	y	3	3
0205	Extract from commercial register	y	n	4	-
0206	Credit information commercial tenants	y	n	4	-
0207	Information provided by tenants	y	n	4	-
0208	Credit assessment, credit report	yn	n	4	-
0209	Overview of rental deposits	y	n	4	-
0210	Deposit	y	n	4	-
0211	Handover protocol	y	n	4	-
0212	Hand-over of keys	y	n	4	-
	Tenant and lease contract information				
0213	Residential lease contract	y	y	3	2
0214	Commercial lease contract	y	y	3	2
0215	Lease agreement, operating lease contract	y	y	3	2
0216	Sublease	y	y	3	2
0217	Parking space lease	y	y	3	2
0218	Lease for advertising space	y	n	4	-
0219	Antenna lease	y	n	4	-
0220	Other license agreements	y	y	3	2
0221	Other agreements	y	n	4	-
0222	Addendum to lease contract	y	n	4	-
0223	Rent guarantee	y	y	3	2
0224	Rent arrears overview	y	n	4	-
0225	Rent arrears tenant	y	n	4	-
0226	Rent adjustment	y	n	4	-
0227	Permanent tenancy rights	y	n	4	-
0228	Rent reduction	y	n	4	-
0229	Exercise of option to extend	y	n	4	-
0230	Correspondence tenant	y	y	3	1
0231	SEPA mandate	y	n	4	-
0232	Notice of defects to tenant	y	n	4	-
0233	Notice to terminate lease	y	n	4	-
0234	Conformation of termination of lease contract	y	n	4	-
0235	Accounting of service charges, utilities	y	n	4	-
0236	Objection to service charge, utilities accounting	y	n	4	-

(continued)

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0237	Property and facility management	y	y	2	2
0238	Management contract overview	y	y	2	2
0239	Management invoice	y	y	2	3
0240	Operating cost overview	y	y	3	1
0241	Property management contract	y	y	3	1
0242	Facilities management contract	y	y	3	1
0243	Center management contract	y	y	3	1
0244	Caretaker contract	y	y	3	1
0245	Building management contract	y	y	3	1
0246	Cleaning service contract	y	y	3	1
0247	Security service contract	y	y	3	1
0248	Snow and ice control contract	y	y	3	1
0249	Power of attorney for property manager	y	n	4	–
0250	Service contract heating	y	y	2	1
0251	Service contract ventilation	y	y	2	1
0252	Service contract air extraction system	y	y	2	1
0253	Service contract air conditioning	y	y	2	1
0254	Service contract sanitary installations	y	y	2	1
0255	Service contract fire protection	y	y	2	1
0256	Service contract fire alarm system	y	y	2	1
0257	Service contract fire damper	y	y	2	1
0258	Service contract fire doors	y	y	2	1
0259	Service contract fire protection gates	y	y	2	1
0260	Service contract hydrants	y	y	2	1
0261	Service contract fire extinguisher	y	y	2	1
0262	Service contract sprinkler system	y	y	2	1
0263	Service contract smoke and heat exhaust ventilation system	y	y	2	1
0264	Service contract escape route control	y	y	2	1
0265	Service contract elevator and escalator	y	y	2	1
0266	Service contract for building management system	y	y	2	1
0267	Service contract for electrical equipment	y	y	2	1
0268	Service contract electro-technical equipment	y	y	2	1
0269	Service contract safety lightning	y	y	2	1
0270	Service contract lightning protection	y	y	2	1
0271	Service contract power generator	y	y	2	1
	Service contract UPS	y	y	2	1

(continued)

Due diligence
in property
transactions

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0272	Service contract emergency power generator	y	y	2	1
0273	Service contract transformer station	y	y	2	1
0274	Service contract barriers	y	y	2	1
0275	Service contract sun protection blinds	y	y	2	1
0276	Service contract doors, roller shutter doors	y	y	2	1
0277	Service contract roof	y	y	2	1
0278	Service contract other	y	y	2	1
	Utilities				
0279	Supply and waste disposal contract	y	y	2	2
0280	Supply contract HVAC	y	y	2	2
0281	Supply contract electricity	y	y	2	2
0282	Supply contract gas	y	y	2	2
0283	Supply contract water	y	y	2	2
0284	Supply contract telecommunications	y	y	2	2
	Insurance				
0285	Insurance overview	y	y	3	2
0286	Third-party liability insurance	y	y	3	1
0287	Fire insurance	y	y	3	1
0288	Terrorism insurance	y	y	3	1
0289	Loss of rent insurance	y	n	3	1
0290	Environmental damage insurance	y	y	3	1
0291	Building insurance	y	y	3	1
0292	Other insurance documents	y	y	3	1
0293	Insured loss	y	y	2	2
0294	Asserted insurance claims	y	y	2	2
	Taxes and public charges				
0295	Land tax assessments, land assessment notice	y	n	4	–
0296	Rating notice, land tax notice	y	n	4	–
0297	Street cleaning charges	y	y	3	2
0298	Surface water drainage, sewage disposal	y	y	3	2
0299	Waste disposal	y	y	3	2
0300	Real estate tax assessment	y	n	4	–
0301	Other public charges	y	n	4	–
0302	Input tax adjustment	y	n	4	–
0303	Corporation tax	y	n	4	–
0304	Value added tax, VAT	y	n	4	–

(continued)

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0305	Value added tax key, VAT key	y	n	4	–
0306	Exemption certificate for construction works	y	n	4	–
0307	Tax certificate	y	n	4	–
0308	Tenant input tax	y	n	4	–
0309	Vendor input tax	y	n	4	–
0310	Option rates, area factor for apportionment	y	n	4	–
0311	Other document				
0312	Litigations	y	n	4	–
0313	Proportion-related work contracts	y	n	4	–
0314	Additional information for specialized properties	y	n	4	–
	Other documents, contracts, agreements	y	n	4	–
	Market and environment				
0315	Market analysis	yn	y	3	1
0316	Location analysis	yn	y	3	1
0317	Infrastructure and transport connections	yn	y	2	3
0318	Analysis of surrounding area	yn	y	3	2
0319	Marketing	y	n	4	–
0320	Website	yn	n	4	–
0321	Press release	y	n	4	–
0322	Tenant's marketing activities	yn	n	4	–
0323	Events	yn	n	4	–
0324	Pictures	n	n	4	–
	Controlling				
0325	Controlling report	y	y	2	2
0326	Outgoing invoices	y	y	3	3
0327	Incoming invoices	y	y	3	3
0328	Building budget control	y	y	2	2
0329	Liquidity calls	y	n	4	–
	Reporting				
0330	Forecast calculation	y	n	4	–
0331	Project updates	y	n	4	–
0332	Monthly reports	y	n	4	–
0333	Quarterly report	y	n	4	–
0334	Financial report	y	n	4	–
0335	Management report	y	n	4	–

(continued)

Due diligence
in property
transactions

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0336	Exit	y	y	1	1
0337	Sale brochure	yn	y	2	3
0338	Sale presentation	y	n	4	—
0339	Agent	y	n	4	—
0340	Sale price estimate	y	n	4	—
0341	Prospective purchasers	y	n	4	—
0342	Letter of intent	y	n	4	—
0343	Sale and purchase agreement (SPA)	y	n	4	—
0344	Q&A	y	n	4	—
	Notarization	y	n	4	—
	<i>Company-related documents</i>				
0345	Company	y	n	4	—
0346	Key data sheet	y	n	4	—
0347	Letter templates	y	n	4	—
0348	Articles of association	y	n	4	—
0349	Commercial register extract	y	n	4	—
0350	Transfer of company shares	y	n	4	—
0351	Register extracts, applications	y	n	4	—
0352	Share register	y	n	4	—
	List of shareholders	y	n	4	—
0353	Special purpose vehicle (SPV)	yn	n	4	—
0354	Structure, organigram	y	n	4	—
0355	Power of attorney	y	n	4	—
0356	Shareholder resolutions	y	n	4	—
0357	Shareholder minutes, minutes of shareholder meeting	y	n	4	—
0358	Annual report	y	n	4	—
	Loan agreements shareholders	y	n	4	—
0359	Constitution	y	n	4	—
0360	Constitution	y	n	4	—
	resolutions to alter constitution	y	n	4	—
0361	Shareholders' resolutions	y	n	4	—
0362	Constitution-related resolutions	y	n	4	—
0363	Profit distribution resolutions	y	n	4	—
0364	Capital distribution resolutions	y	n	4	—
	Other resolutions	y	n	4	—

(continued)

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0365	Management	y	n	4	-
0366	Resolutions of the managing directors	y	n	4	-
0367	Minutes, records	y	n	4	-
0368	Agreements with management				
0369	Company-related contracts				
0370	Accounting	y	n	4	-
0371	Auditors	y	n	4	-
0372	Tax consultants	y	n	4	-
0373	Other contracts	y	n	4	-
0374	Finances				
0375	Taxes	y	n	4	-
0376	Management reporting	y	n	4	-
0377	Liabilities (incl. off-balance sheet)	y	n	4	-
0378	Current balances	y	n	4	-
0379	Other financial information	y	n	4	-
0380	Interim statements (balance sheet and income statement)	y	n	4	-
0381	Fund fees	y	n	4	-
0382	Joint ventures				
0383	Company concept	yn	n	4	-
0384	Legal form	yn	n	4	-
0385	Cooperation agreement	y	n	4	-
0386	Capital participation	y	n	4	-
0387	Minutes, records	y	n	4	-
0388	Other documents	y	n	4	-
0389	Financing				
0390	Bank address	y	n	4	-
0391	Term sheet	y	n	4	-
0392	Loan agreement	y	n	4	-
0393	Condition agreements	y	n	4	-
0394	Guarantees	y	n	4	-
0395	Equity capital	y	n	4	-
0396	Debt capital	y	n	4	-
0397	Interest and amortization notices	y	n	4	-
0398	Disbursement requests	y	n	4	-
0399	Securities	y	n	4	-
0400	Interest rate hedge	y	n	4	-

(continued)

Due diligence
in property
transactions

Class label	Document classes	Text-based classification	Subject of TDD	Priority	Machine readability score
0396	Valuation reports	y	n	4	-
0397	Deal announcement	y	n	4	-
0398	Power of attorney	y	n	4	-
0399	Current bank accounts	y	n	4	-
0400	Account management	y	n	4	-
0401	Account authorization	y	n	4	-
	Funds				
0402	Fund documents	y	n	4	-
0403	Fund controlling	y	n	4	-
0404	Committee template	y	n	4	-
0405	Depository	y	n	4	-
0406	Fund approval process	y	n	4	-
0407	Liquidity management	y	n	4	-
0408	Risk management	y	n	4	-
0409	Fund accounting	y	n	4	-
0410	Sales support	y	n	4	-