System Design Specification

Table of Contents

1. Introduction

To process personal data under GDPR, companies are required to transparently communicate to their end users **which legal ground** (e.g., consent; legitimate interest; etc.) applies to the collection of a certain data point. Depending on that legal ground, different individual rights apply or don't apply. Propose a way to **automatically analyze privacy policies** of online services to extract the legal ground that they claim to use for a specific subset of data, propose a way of how this information could be **visualized** to an end user. Finally, based on the proposed way of extracting this information, design a system that **automatically acts on specific rights** (e.g., erasure) on behalf of the user (data subject)?

2. Purpose

The purpose of this document is to provide a comprehensive overview of the system design for the **automated analysis of privacy policies** of online services and **the automated exercise of data subject rights**. It outlines the architecture, key functional components, user interfaces, and integration aspects of the system.

3. System Overview

System Architecture

The system follows a client-server architecture. The client side includes the user interface module, while the server side consists of the backend modules responsible for privacy policy extraction, text processing, legal ground identification (traditional text parser and a 3rd party NLP service), rights mapping(GDPR), user selections, and automated actions (data subject legal rights).

Key Components

1. An Interactive User Interface:
   a) Input – user identification, e.g. user account, password, email account (if necessary), etc.
   b) Input – user consent
   c) Input – target web services URL to be analyzed for legal ground extraction
   d) Input – user select actions: the actions taken by the user within the system to indicate their choices and preferences regarding the exercise of their data subject rights
   e) Output – implemented system's privacy policy;
   f) Output – extracted legal grounds of 3rd party web services, and related commercial explanation and communication to consumers;
   g) Output – legal rights associated to each identified legal grounds
2. User Information and Consent: Manages user consent and presents information about the system's privacy policy and legal grounds.
3. Privacy Policy Extraction: Extracts privacy policy text from online services using web scraping

techniques.

4. Text Processing and Legal Ground Identification: Utilizes natural language processing techniques to analyze privacy policy text and identify legal grounds associated with different data points.
5. Rights Mapping: Maps identified legal grounds to corresponding individual rights under the GDPR. (Verification by legal professionals are required)
6. User Selections and Data Subject Actions: Allows users to select specific rights they wish to exercise and facilitates data subject actions.
7. Automated Action on Behalf of Data Subject: Generates and performs automated actions (e.g., erasure requests) on behalf of the data subject to relevant online services.

An expansion of the solution consists of transforming the solution into a browser extension that functions like an Ad-blocker. The extension will be able to autonomously search, retrieve and analyze privacy policies, while also enabling data subjects to exercise their rights. This design will eliminate the manual entry of third-party website URLs and enhance the user experience by seamlessly embedding the functionality into the browsing activity itself.

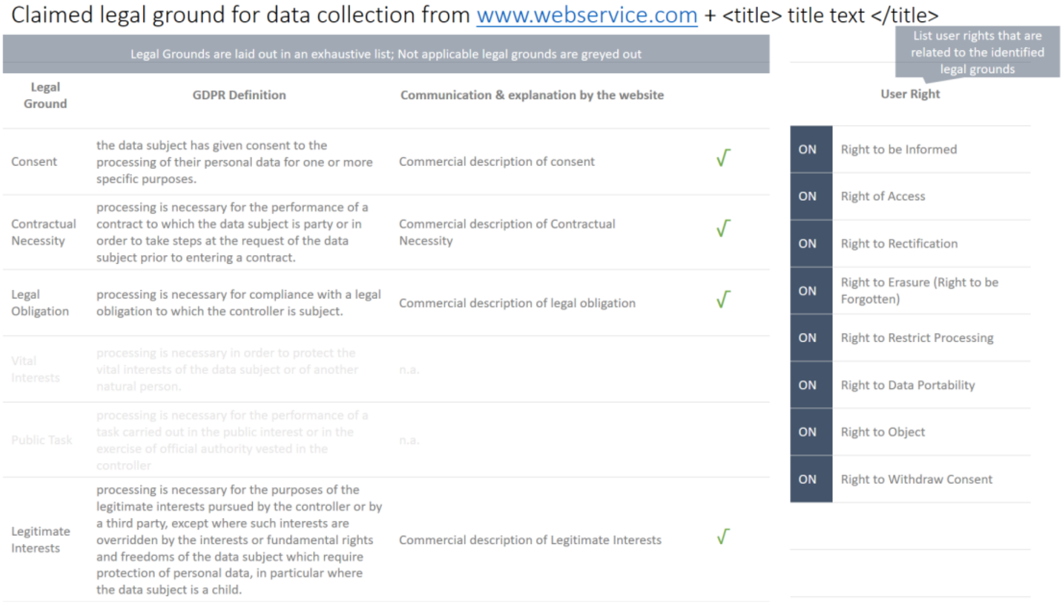
Integration with External Services

The system shall integrate with external services through API calls, if defined and provided, to send automated action instructions such as erasure requests, or other means of communication required by the external service provider. It is expected that *ad hoc* protocols and APIs will be negotiated and agreed with external service providers. It must ensure secure communication and adherence to the specific protocols and formats required by each service.

4. Functional Requirements

User Interface

- Displays own privacy policy and acquire user consent and capture user privacy preferences to use the functionality of this system
- Allows users to input the name or URL of an online service's privacy policy. (this UI functionality becomes obsolete in the browser-extension version of the design)
- Displays the extracted legal grounds with respective legal definitions (by GDPR), the commercial explanation of the service provider, and the applicable rights of the user.
- Enables users to select specific rights they wish to exercise and related data collection to execute the user specified right(s), e.g. 3rd party user ID, required credentials, etc.



Claimed legal ground for data collection from www.webservice.com + <title> title text </title>

Legal Grounds are laid out in an exhaustive list; Not applicable legal grounds are greyed out

Legal Ground	GDPR Definition	Communication & explanation by the website	
Consent	the data subject has given consent to the processing of their personal data for one or more specific purposes.	Commercial description of consent	√
Contractual Necessity	processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering a contract.	Commercial description of Contractual Necessity	√
Legal Obligation	processing is necessary for compliance with a legal obligation to which the controller is subject.	Commercial description of legal obligation	√
Vital Interests	processing is necessary in order to protect the vital interests of the data subject or of another natural person.	n.a.	
Public Task	processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller	n.a.	
Legitimate Interests	processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.	Commercial description of Legitimate Interests	√

List user rights that are related to the identified legal grounds

	User Right
ON	Right to be Informed
ON	Right of Access
ON	Right to Rectification
ON	Right to Erasure (Right to be Forgotten)
ON	Right to Restrict Processing
ON	Right to Data Portability
ON	Right to Object
ON	Right to Withdraw Consent

User Information and Consent

- Presents users with information about the system's privacy policy, legal grounds, and data processing.
- Obtains explicit consent from users to process their data. Enable/disable remote functionalities based on users’ consent choices.

Privacy Policy Extraction

- Uses web scraping techniques to navigate websites and extract privacy policy and cookies policy texts. Websites use cookies to obtain user data, therefore, cookies policy shall be scrapped and analyzed together with privacy policy.
- Stores the extracted privacy policy and cookies text in a (local) database or data storage for further processing.

Text Processing and Legal Ground Identification

- Applies natural language processing techniques to analyze the extracted privacy policy text.
  - ■ Local processing
  - ■ Remote processing/3rd NLP services
- Identifies key phrases related to legal grounds in the privacy policy text.
- Trains a machine learning model to classify legal grounds associated with different data points.

Specifically, develop a list of keywords or phrases associated with different legal grounds (e.g., "consent," "contractual necessity," "legal obligation," etc.). Use keyword matching or regular expressions to identify instances of these keywords in the processed privacy and cookies policy texts:

General Data Protection Regulation (GDPR) defined 6 types of “legal ground” for processing personal data. These legal grounds provide the basis for lawful processing of personal data.

1. Consent (Article 6(1)(a)):

- The data subject has given clear and explicit consent for the processing of their personal data for a specific purpose.

2. Contractual Necessity (Article 6(1)(b)):

- The processing is necessary for the performance of a contract to which the data subject is a party, or to take steps at the request of the data subject prior to entering into a contract.

3. Legal Obligation (Article 6(1)(c)):

- The processing is necessary for compliance with a legal obligation to which the data controller is subject.

4. Vital Interests (Article 6(1)(d)):

- The processing is necessary to protect the vital interests of the data subject or another natural person.

5. Public Task (Article 6(1)(e)):

- The processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the data controller.

6. Legitimate Interests (Article 6(1)(f)):

- The processing is necessary for the legitimate interests pursued by the data controller or a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject.

<table>
<tr><td>Step 1:</td><td colspan="3">1-1)Use key word “Privacy”, “Privacy Policy”, “Cookies Policy” on homepage to locate and retrieve the full text of website’s policy;
1-2)Locate and retrieve the full text of cookie notifications or consent pop-ups</td></tr>
<tr><td rowspan="5">Step 2:</td><td colspan="3">Parse the text and look for legal ground key words, which could be positively identified by looking up for the following:</td></tr>
<tr><td>Direct text reference (case-insensitive)</td><td>Legal clause reference (case-insensitive)</td><td>Implicit reference (Fuzzy search)</td></tr>
<tr><td><b>Consent</b></td><td><b>Article 6(1)(a)</b>
<b>Art. 6(1)(a)</b>
<b>Article 6-1-a and alike</b></td><td></td></tr>
<tr><td><b>Contractual Necessity</b></td><td><b>Article 6(1)(b)</b>
<b>Art. 6(1)(b)</b>
<b>Article 6-1-b and alike</b></td><td></td></tr>
<tr><td><b>Legal Obligation</b></td><td><b>Article 6(1)(c)</b></td><td></td></tr>
</table>

		Art. 6(1)(c) **Article 6-1-c and alike**	
	Vital Interests	**Article 6(1)(d)** **Art. 6(1)(d)** **Article 6-1-d and alike**	
	Public Task	**Article 6(1)(e)** **Art. 6(1)(e)** **Article 6-1-e and alike**	
	Legitimate Interests	**Article 6(1)(f)** **Art. 6(1)(f)** **Article 6-1-f and alike**	

An intermediate text file containing the content of the privacy/cookies policy (notice) of the target website is created for further processing. The file is constructed in a machine-readable format (key value pair), and contains nature language content, so that it can be fed directly to a trained (3rd party) NLP engine. In the prototype, claimed legal ground(s) are identified by means of a (local) traditional fuzzy matching algorithm as well as by invoking remote third-party (OpenAI) NLP analysis.

Rights Mapping

- Establishes a mapping between identified legal grounds and the corresponding individual rights under the GDPR.
- Defines rules to determine which rights apply based on the identified legal ground.

Legal Ground defined by GDPR	Service provider explanation	GDPR defined user right
Consent: the data subject has given consent to the processing of their personal data for one or more specific purposes;		- Right to withdraw consent - Right to be informed - Right to access personal data - Right to rectification - Right to erasure (right to be forgotten) - Right to restrict processing - Right to data portability
Contractual Necessity processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the		- Right to access personal data - Right to rectification - Right to restrict processing

request of the data subject prior to entering into a contract		
Legal Obligation processing is necessary for compliance with a legal obligation to which the controller is subject		- Right to access personal data - Right to rectification - Right to erasure (right to be forgotten) - Right to restrict processing - Right to data portability
Vital Interests processing is necessary in order to protect the vital interests of the data subject or of another natural person		- Right to access personal data - Right to rectification - Right to erasure (right to be forgotten) - Right to restrict processing - Right to data portability
Public Task processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller		- Right to access personal data - Right to rectification - Right to erasure (right to be forgotten) - Right to restrict processing - Right to data portability
Legitimate Interests processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child		- Right to object to processing - Right to restrict processing

Reversely, legal rights can be mapped towards legal grounds as follows:

GDPR Legal Right	GDPR Legal Grounds

Right to be Informed	● Consent ● Contractual necessity ● Legal obligation ● Legitimate interests
Right of Access	● Consent ● Contractual necessity ● Legal obligation ● Legitimate interests
Right to Rectification	● Consent ● Contractual necessity ● Legal obligation ● Legitimate interests
Right to Erasure (Right to be Forgotten)	● Consent (withdrawn) ● Data no longer necessary, Processing is unlawful, Legal obligation, Legitimate interests
Right to Restrict Processing	● Accuracy disputed, Processing is unlawful, Data no longer necessary, Legal claims, Legitimate interests
Right to Data Portability	Consent, Contractual necessity
Right to Object	Legitimate interests, Public task or official authority, Direct marketing
Rights Related to Automated Decision Making and Profiling	Explicit consent, Contractual necessity, Legal obligation, Legitimate interests
Right to Withdraw Consent	Consent

Right to Lodge a Complaint	N/A

User Selections and Data Subject Actions

- Presents the extracted legal grounds with their respective legal definition and the service provider’s commercial explanation
- Present applicable rights associated to the identified legal grounds to the user.
- Allows users to select specific rights they wish to exercise.
- Captures user selections and preferences regarding the specific actions they want to perform.

Automated Action on Behalf of Data Subject

- Generates standard requests (e.g., erasure requests) based on user selections.

The system shall integrate with external services through API calls, if defined and provided by the service provider, to send automated action instructions such as an ***erasure*** request, or other means of communication required by the external service provider. It is expected that *ad hoc* protocols and APIs will be negotiated and agreed with external service providers.

- Automatically fill in necessary details (e.g., user's personal information, data identifiers) in request templates.
- When user defined/selected trigger condition(s) are met, send the generated requests securely to the relevant online service provider on behalf of the data subject.

5. Non-Functional Requirements

Security and Data Protection

- The system shall ensure the secure handling and storage of user data, following industry-standard encryption and security practices.
- User data shall be anonymized and pseudonymized whenever possible to protect privacy.
- Access controls and authentication mechanisms shall be implemented to restrict unauthorized access to user data.

Performance

- The system shall be designed to handle a high volume of privacy policy analysis and user requests efficiently.
- The system shall be designed to handle anti-crawling mechanisms of the 3rd party service

provider and prevent hanging.
- Response times for extracting privacy policies, analyzing legal grounds, and generating automated actions shall meet acceptable performance standards.

Scalability, Reliability and Maintainability

Given the nature of the task, there is no consideration on scalability, reliability and maintainability during the current phase of design.

6. System Integration

APIs and Services

- The system shall integrate with external services through well-defined APIs to send automated actions.
- When and if APIs to retrieve privacy policies and cookies policy are provided, it can be directly used instead of scrapping the full website.

Data Storage

- Privacy policy texts and associated metadata shall be stored in a database or data storage system.
- Proper data indexing and retrieval mechanisms shall be implemented to enable efficient search and retrieval.

User Authentication and Authorization

- User authentication and authorization mechanisms shall be implemented to ensure that only authorized users can access and interact with the system.
- User credentials and access privileges shall be securely managed and stored.

7. System Deployment

Deployment Architecture

- The current design is a local application with remote data processing functionality, the remote system shall be deployed (in the future) in a scalable and redundant architecture to ensure high availability and fault tolerance. Load balancers, application servers, and database servers shall be set up to handle user requests and manage data.
- In the browser-extension version, publish the browser-extension with mainstream webbrowsers.

Hardware and Software Requirements

- Hardware requirements shall be determined based on anticipated system load and scalability requirements.
- If the system is planned for commercial deployment, it shall be designed to be platform-independent, supporting various operating systems and databases.

8. Testing Strategy

Unit Testing

- Each component and module shall be thoroughly tested in isolation to ensure their individual functionality.
- Test cases shall be designed to cover different scenarios, including edge cases and error conditions.

Integration Testing

- Integration tests shall be conducted to verify the interaction and compatibility of system components.
- API integrations with external services shall be tested to ensure proper communication and data exchange.

User Acceptance Testing

- User acceptance tests shall be conducted to validate the system's functionality, usability, and compliance with user requirements.
- Realistic test scenarios shall be executed to simulate user interactions and actions.

Privacy and Compliance

Data Handling and Storage

- The system itself shall handle personal data in accordance with applicable data protection laws, including the GDPR.
- Personal data shall be stored securely, with appropriate access controls and encryption mechanisms in place.
- The system shall adhere to GDPR principles, including the lawful basis for processing personal data and respecting individual rights.

- User consent management shall be implemented to record and manage user consent preferences.

9. Monitoring and Maintenance

Logging and Error Handling

- The system shall log relevant events, errors, and exceptions for monitoring and troubleshooting purposes.
- Proper error handling mechanisms shall be implemented to gracefully handle errors and exceptions.

Maintenance Processes

- Regular maintenance activities, including updates, bug fixes, and security patches, shall be planned and executed.
- Change management processes shall be followed to minimize disruption during maintenance activities.

10. Conclusion

This system design document provides a comprehensive overview of the proposed solution for automated privacy policy analysis and the exercise of data subject rights. It outlines the system architecture, key components, integration aspects, functional and non-functional requirements, as well as the testing, privacy, and maintenance strategies. By following this design, the system can effectively and efficiently handle privacy policy analysis, empower users to exercise their rights, and ensure compliance with data protection regulations.

11. Appendices

Glossary

- A list of relevant terms and definitions used throughout the document.

1. GDPR: General Data Protection Regulation, a regulation in the European Union that aims to protect the personal data and privacy of individuals within the EU.
2. Privacy Policy: A document that outlines how an organization collects, uses, stores, and protects the personal data of its users or customers.
3. Legal Ground: The legal basis or justification for the processing of personal data, as defined by the GDPR.
4. User Interface: The interface through which users interact with the system, providing input and receiving output.

5. Web Scraping: The process of extracting data from websites by automatically navigating through web pages and collecting relevant information.
6. Natural Language Processing (NLP): A field of artificial intelligence that focuses on the interaction between computers and human language, enabling computers to understand and analyze natural language text.
7. API: Application Programming Interface, a set of rules and protocols that allows different software applications to communicate and interact with each other.
8. User Authentication: The process of verifying the identity of a user, usually through credentials such as username and password.
9. User Authorization: The process of granting specific access rights and permissions to users based on their authenticated identity.
10. Data Storage: The mechanism used to store and manage data, which can include databases, file systems, or cloud storage.
11. User Acceptance Testing: Testing performed by end users to ensure that the system meets their requirements and is usable.
12. Encryption: The process of encoding data in such a way that only authorized parties can access and understand it.
13. Anonymization: The process of removing or encrypting personally identifiable information from data to prevent the identification of individuals.
14. Pseudonymization: The process of replacing personally identifiable information with pseudonyms, which can still be used for analysis but does not directly identify individuals.
15. Load Balancer: A device or software component that distributes network traffic across multiple servers to ensure efficient utilization and high availability.
16. Change Management: A structured approach to manage changes to the system, including updates, bug fixes, and security patches, while minimizing disruption to users.

Functional Chart: