

## KNOWLEDGE-ORIENTED SEARCH ENGINE ALGORITHM

The main objective is to prepare a proposition of a technical solution and a substantive concept for an internet site operating as an information hub for the publicly available datasets of the Institute of National Remembrance. We wish to develop a solution serving as an integrative platform with a multi-search engine for the public datasets shared online by the INR on the one hand, while on the other, to serve as a research platform for both the users of the solution as well as for the INR.

Together with you, we wish to create an advance knowledge base related to the modern history of Poland, used by pupils, teachers, students, as well everyone who is passionate about history.

The solution must be intuitive, easy to use and contain creative solutions regarding both the interface, as well as for filtering expressions and data. The project is dedicated to the representatives of the z generation and millennials, but its transparency should also make it interesting for the so-called boomers. There should be an option to export the search results to an external file. The solution must have an option to introduce a list of input data in the form of a file or a form, as well as the possibility of verifying the results based on the indicated sources of information.

The project must be prepared simultaneously in at least two language versions - Polish and English and adapted to the WCAG 2.1. requirements. The solution must be delivered as a clickable mock-up.

Preparing a creative name for the project also constitutes a part of the task.

### **I) Minimum technical parameters of the search engine:**

1. The engine indexes objects located on the INR website (using a list programmed in the administration panel), including text attachments, pdf, graphics, multimedia, pages and subpages (except for technical pages, such as print), in total over 10,000,000 objects;
2. The engine has the capacity to index the INR databases;
3. Databases are indexed automatically, without the need to use an external, non-standard connector (so-called custom-built);
4. Minimum search efficiency at 30 queries per second for a set of 500,000 objects;
5. Response time for 80% queries below 1 second for the standard user interface with at least 30 parallel queries for an index of 500,000 objects in accordance with the device's log;

### **II) Minimum functional requirements of the administration panel:**

1. Graphical interface of the administration console with all configuration options;
2. User interface definable from the administration console.
3. Administration panel allows for the possibility of controlling system functionalities.
4. The possibility of adding subsequent websites and databases for indexing;
5. Access to standard search logs;
6. Administration events view in logs;

### **III) Minimum functional requirements of the search engine:**

1. Polish language support, including inflection;
2. Indexing all videos from the IPNtvPL channel on YouTube;
3. Metadata analysis for the purpose of building navigation in search results;
4. Mechanism of automated query optimisation based on user choices (clicks) of specific query results;
5. Generating preview of indexed office documents (in MS Office, PDF format) through "Preview" button"
6. Dictionary of synonyms including the possibility for the administrator to define custom terms;
7. The ability to promote the results by website source address.
8. The ability to select the number of search results displayed on the website (by the search engine user).
9. Full-text search mechanism, including in the contents of the following formats of files posted on the Orderer's website: .pdf, native MS Word formats
10. The search engine proposes a suggestion for searching another expression or a phrase in cases where the user searches for expressions and phrases which generate no results. Presenting results ranked by popularity - the initial 10 positions.
11. If no results are found, the search engine returns expressions based on user profile and expressions recommended by the administrator.
12. The option to include abbreviations in the search results, such as: AK – Armia Krajowa
13. The ability to count the search results by the external statistics registration systems;
14. The ability to search using all indexes or selectively selected from the defined list. A field to declare the website or list of websites declared for searching using a multiple-choice list.
15. Displaying search results in the following manner:

- By default per search, results sorted by search accuracy taking into account popularity statistics for search results and selecting from lists of results for previous searches;
- By default for searching using the search field located on other websites of the Orderer, results sorted taking source into account, while the search results from the website used to make the search rank highest, followed by results from other websites per search accuracy taking into account popularity statistics for results and selections from search result lists for previous searches;
- The results can be sorted by defining priorities for the following conditions:
  - search accuracy - default search
  - from newest to oldest and in reverse order - if mutually defined tags are placed, allowing to prepare the functionality.
- For the searched articles, the search result should present (if such information is available in the indexed set) the number of found objects, category, titles, publication dates, and in the case of full-text search, a paged list featuring a fragment of 240 characters including spaces of the article containing the searched phrase or the article header;
- The results should be presented in full-text method, taking into account the &lt;title&gt;, &lt;lead&gt;, &lt;text&gt; fields.
- The results can be filtered by place of publication, based on the returned results after searching in all indexes.

#### IV) Minimum user profiling parameters:

1. Automatic knowledge base definition using search history;
2. Classification and aggregation of users and data based on input criteria (i.e. age groups, interests, ...);
3. Using a defined knowledge base in the search results and user suggestions.

#### A list of websites of the Institute of National Remembrance (Central Office and Branch Offices)

No.		Website name	Internet address
1		INR main site	<a href="https://ipn.gov.pl">https://ipn.gov.pl</a>
2		INR main site in English	<a href="https://ipn.gov.pl/en/">https://ipn.gov.pl/en/</a>
3		INR Public Information Bulletin	<a href="https://bip.ipn.gov.pl/">https://bip.ipn.gov.pl/</a>
4	EDUCATIONAL	Education	<a href="https://edukacja.ipn.gov.pl/">https://edukacja.ipn.gov.pl/</a>
5		Biographies of Polish Heroes	<a href="https://biogramy.ipn.gov.pl/">https://biogramy.ipn.gov.pl/</a>
6		Giants of Science	<a href="https://gigancinauki.pl/">https://gigancinauki.pl/</a>
7		Katyn 1940	<a href="https://katyn.ipn.gov.pl/">https://katyn.ipn.gov.pl/</a>
8		Border Protection Corps	<a href="https://kop.ipn.gov.pl/">https://kop.ipn.gov.pl/</a>
9		Independent Culture	<a href="https://kultura-niezalezna.pl/">https://kultura-niezalezna.pl/</a>
10		Remembrance Maps	<a href="https://mapypamieci.ipn.gov.pl/">https://mapypamieci.ipn.gov.pl/</a>

11		My Independent	<a href="https://mojaniepodlegla.pl/">https://mojaniepodlegla.pl/</a>
12		Stories Told	<a href="https://opowiedziane.ipn.gov.pl/">https://opowiedziane.ipn.gov.pl/</a>
13		Papal Pilgrimages in the Polish People's Republic Documented	<a href="https://pielgrzymki.ipn.gov.pl/">https://pielgrzymki.ipn.gov.pl/</a>
14		Polish Months	<a href="https://polskiemiesiace.ipn.gov.pl/">https://polskiemiesiace.ipn.gov.pl/</a>
15		Greater Poland Uprising	<a href="https://pw.ipn.gov.pl/">https://pw.ipn.gov.pl/</a>
16		Next Stop History	<a href="https://przystanekhistoria.pl/">https://przystanekhistoria.pl/</a>
17		Flighting Solidarity	<a href="https://sw.ipn.gov.pl/">https://sw.ipn.gov.pl/</a>
18		Truth About Camps	<a href="https://truthaboutcamps.eu/">https://truthaboutcamps.eu/</a>
19		Polish September 1939	<a href="https://1wrzesnia39.pl/">https://1wrzesnia39.pl/</a>
20		Free Trade Unions	<a href="https://wzz.ipn.gov.pl/">https://wzz.ipn.gov.pl/</a>
21		Volhynia Massacre	<a href="https://zbrodniawolynska.pl/">https://zbrodniawolynska.pl/</a>
22		Cursed Soldiers – Forgotten Heroes	<a href="https://podziemiezbrojne.ipn.gov.pl/">https://podziemiezbrojne.ipn.gov.pl/</a>
23		Life for a Life	<a href="https://zyciezazycie.pl/">https://zyciezazycie.pl/</a>
24		THE “POLISH OPERATION” OF THE NKVD 1937–1938	<a href="https://operacja-polska.pl/">https://operacja-polska.pl/</a>
25		Martyrdom of Polish Villages	<a href="https://martyrologiawspolskich.pl">https://martyrologiawspolskich.pl</a>
26		“Wretched moral midgets of the reaction...” Polish Independence Underground 1944–1956	<a href="https://zkr.ipn.gov.pl/">https://zkr.ipn.gov.pl/</a>
28		Pomeranian Crime	<a href="https://zbrodniapomorska1939.pl/">https://zbrodniapomorska1939.pl/</a>
29		From independence to independence, History of Poland 1918–1989	<a href="https://www.polska1918-89.pl/">https://www.polska1918-89.pl/</a>
30	INSTITUTIONAL	The Archive Full of Remembrance	<a href="https://archiwumpamieci.pl/">https://archiwumpamieci.pl/</a>
31		Bureau of Searches and Identification	<a href="https://poszukiwania.ipn.gov.pl/">https://poszukiwania.ipn.gov.pl/</a>
32		Next Stop History Educational Centre	<a href="https://centrumedu.ipn.gov.pl/">https://centrumedu.ipn.gov.pl/</a>
33		Centre of Information on the Victims of World War II	<a href="https://ofiary.ipn.gov.pl/">https://ofiary.ipn.gov.pl/</a>
34		The Cross of Freedom and Solidarity	<a href="https://ipn.gov.pl/kwi">https://ipn.gov.pl/kwi</a>
35		Opposition	<a href="https://opozycja.ipn.gov.pl/">https://opozycja.ipn.gov.pl/</a>
36		INR Archival Review	<a href="https://ipn.gov.pl/par/">https://ipn.gov.pl/par/</a>
37		Poland In The Heart Of European History	<a href="https://pheh.pl/">https://pheh.pl/</a>
38		The European Network of Official Authorities in Charge of the Secret-Police Files	<a href="https://sovietblocksecurity.eu">https://sovietblocksecurity.eu</a>
39		Strzelecka 8 – in development	<a href="https://strzelecka8.ipn.gov.pl/">https://strzelecka8.ipn.gov.pl/</a>
40	SPECIALISED	Bibliography of Polish History	<a href="https://bibliografia.ipn.gov.pl/">https://bibliografia.ipn.gov.pl/</a>
41		Encyclopaedia of Solidarity	<a href="https://encysol.pl/">https://encysol.pl/</a>
42		The Archival Inventory	<a href="https://inwentarz.ipn.gov.pl/">https://inwentarz.ipn.gov.pl/</a>
43		Screening Office Catalogues	<a href="https://katalog.bip.ipn.gov.pl/">https://katalog.bip.ipn.gov.pl/</a>
44		Index of the Repressed	<a href="https://indeksrepresjonowanych.pl/">https://indeksrepresjonowanych.pl/</a>
45		Screening Declaration Information System	<a href="https://data-oswiadczenia.ipn.gov.pl/">https://data-oswiadczenia.ipn.gov.pl/</a>
46	COMPETITIONS	Undeclared 1918-2018	<a href="https://niezwyciezeni1918-2018.pl/">https://niezwyciezeni1918-2018.pl/</a>
47		Policemen in Service of History	<a href="https://policjanciwhistorii.ipn.gov.pl/">https://policjanciwhistorii.ipn.gov.pl/</a>
48		Soldiers in Service of History	<a href="https://zolnierzewhistorii.ipn.gov.pl/">https://zolnierzewhistorii.ipn.gov.pl/</a>
49		“Echoes of Katyn” International Film Festival On Totalitarianisms	<a href="https://totalitaryzmy.pl/">https://totalitaryzmy.pl/</a>