

Product Requirements Document (PRD)

Project: Intention-Ally Search Clustering & Visualization

Version 1.0

Date: March 2025

Owner: [Your Name]

1. Overview

1.1 Product Summary

Intention-Ally is a semantic search and clustering tool that organizes search results into an interactive mind map. It helps users navigate complex topics like textile sustainability, digital product passports, and the future of fashion by clustering related articles based on contextual similarity and trustworthiness.

The tool integrates Brave Search API (and optionally Tavily API) to retrieve high-quality search results, then applies NLP clustering techniques (UMAP + HDBSCAN) to group related articles dynamically. Users can interact with the results via a D3.js-based mind map, filter results by trust score, and optionally save interaction history to refine future searches.

2. Objectives & Goals

2.1 Business Goals

- Improve information discovery for professionals researching sustainability and fashion-tech.
- Provide an intuitive mind map UI to help users explore search results contextually.
- Allow users to filter results by reputation, topic density, and time.
- Enhance personalization by enabling users to rank sources and store search history.

2.2 Technical Goals

- Use Brave API for search, with optional Tavily API for additional metadata.
- Implement NLP-based clustering to group similar articles.
- Use D3.js for visualizing results as an interactive graph.
- Implement trust-weighted scoring to prioritize high-quality sources.
- Allow users to save their interactions to improve future clustering.

3. Key Features & Functionality

3.1 Data Retrieval

- Query Brave API for search results (titles, URLs, snippets).

- Optional: Use Tavily API for entity extraction and additional metadata.
- Store retrieved data in JSON or SQLite.

3.2 NLP-Based Clustering

- Convert article text into vector embeddings (Sentence Transformers).
- Reduce dimensionality using UMAP.
- Apply HDBSCAN clustering to group semantically similar articles.

3.3 Trust-Based Ranking

- Allow users to rate sources (trusted, neutral, avoid).
- Assign a trust score to each cluster based on user inputs.
- Downrank clusters containing low-trust sources.

3.4 Search Filtering Per Cluster

- Filter search results based on:
 - * Minimum reputation score
 - * Topic match strength (e.g., >80% similarity)
 - * Time range (e.g., recent articles only)

3.5 UI: D3.js-Based Graph Mind Map

- Node size = Article relevance/trust score.
- Edges = Contextual similarity between articles.
- Hovering over a node = Preview article metadata.
- Clicking a node = Expand full article details.

3.6 User Interaction History (Optional)

- Allow users to save their interactions to refine future searches.
- Use past clicks, trust ratings, and search behavior to improve clustering.

4. Technical Stack & Architecture

Backend Processing:

- Search API: Brave API, Tavily API (optional)
- NLP Clustering: UMAP + HDBSCAN + Sentence Transformers
- Trust Ranking: User-Weighted Scores
- Database: SQLite / JSON
- Backend: FastAPI (Python)

Frontend UI:

- Graph Visualization: D3.js
- Backend API: FastAPI
- Hosting: Replit

5. User Flow & Experience

User enters a query -> Search results retrieved -> NLP clusters similar articles -> Trust-weighted ranking refines results ->

D3.js graph displays clusters -> User explores results -> User optionally saves interaction history.

6. Open Questions

1. How should we store user trust ratings? (SQLite, JSON, or cloud-based?)
2. Do we need a user authentication system? (for saving history)
3. Should we provide a text-based alternative to the mind map UI?

7. Future Enhancements (Post MVP)

- Multi-language support
- User profiles & custom dashboards
- Real-time collaboration features

8. Timeline & Milestones

Week 1-2: API Integration (Brave, Tavily)

Week 3: NLP Clustering (UMAP + HDBSCAN)

Week 4-5: Trust-Based Ranking & Filters

Week 6: D3.js UI for Graph Visualization

Week 7: User Interaction History (Optional)

Week 8: Testing & Deployment on Replit

9. Conclusion

This project will revolutionize search exploration by moving beyond linear search results and into intelligent, context-driven clustering.