

TinyWorld AI – Idea Summarizer

*Fully offline, lightweight (<10 MB) text
summarization system.*



Presented By:
**Samaksh
Parihar**

Problem Statement

Why Text Summarization?

Modern users deal with:

Long articles

Reports

Stories

Technical documents



Goal:

Create a system that:

Works offline

Is lightweight (<10 MB)

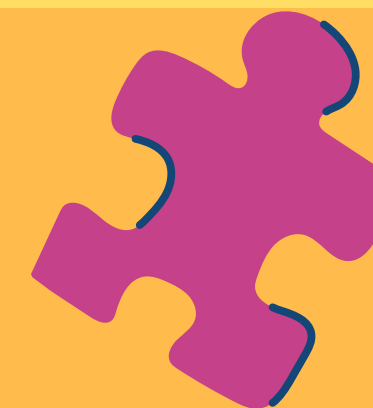
Produces human-like summaries

Lets the user choose summary length

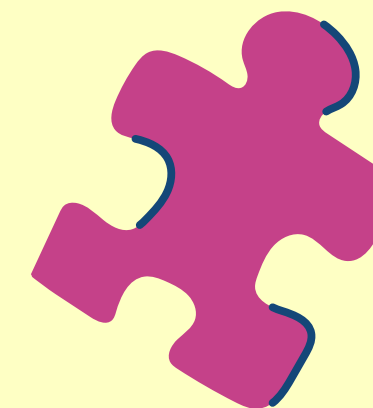
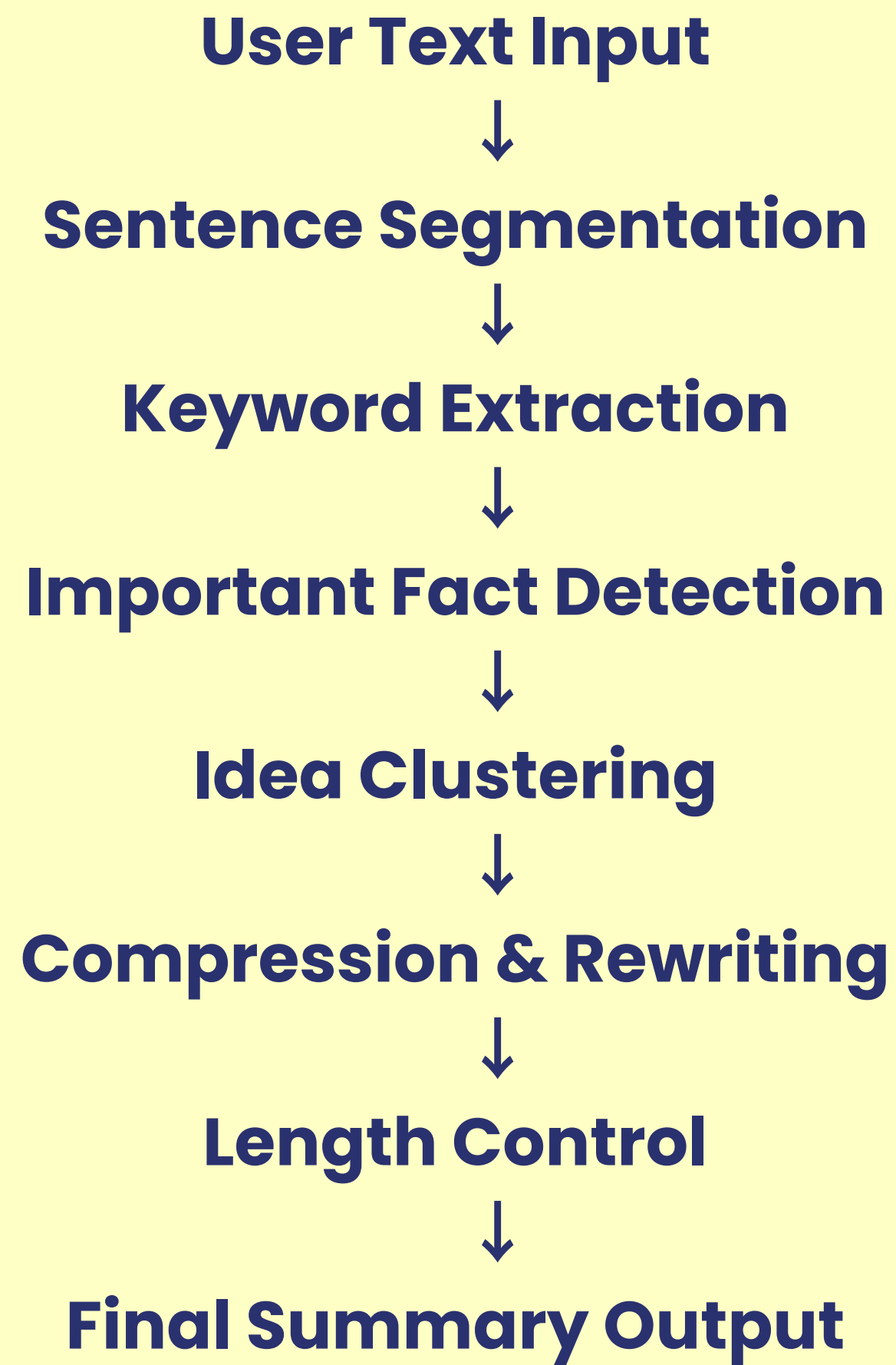
Works for any kind of text

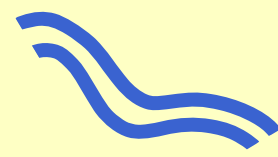
✓ My Solution

TinyWorld AI – Idea Summarizer follows an idea-centric approach rather than relying on sentence extraction. The system first splits the input text into sentences, then identifies keywords that represent the core concepts. It selects sentences that contain multiple keywords to detect important facts. These facts are then clustered into groups based on similarity, compressed into concise statements, and assembled into a fluent, human-like summary. Users can control the length of the summary by choosing short, medium, or long options, making the system flexible and adaptive.



✓ My Solution





Step 1 – Sentence Segmentation

What happens?

The input text is split into individual sentences.

Why?

Humans understand text as ideas expressed in sentences.

How?

Input Paragraph → List of Sentences

Step 2 – Keyword Extraction

Method

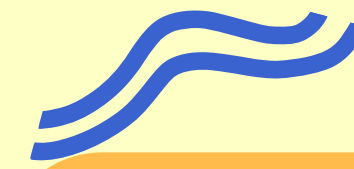
- Extract meaningful words (length ≥ 4)
- Remove stopwords (the, is, was, etc.)
- Count word frequency
- Keep top keywords

Example

Text about electric cars →

Keywords:

electric, battery, charging, technology, range



Step 3 – Important Fact Detection

Approach

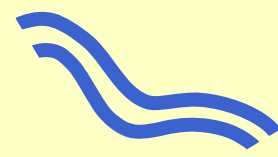
Each sentence is scored:

- +1 point per keyword found

Only sentences with multiple keywords are kept.

Why?

- Removes fluff
- Keeps only information-dense sentences



Step 4 – Idea Clustering

What happens?

Instead of selecting sentences randomly, we:

- Group sentences that talk about the same idea

Example

Cluster 1:

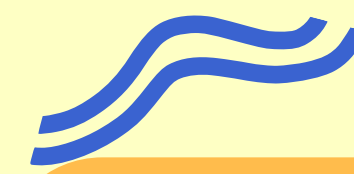
- Battery improvements
- Faster charging

by this the model focus on the idea of the text written

Step 6 – Length Control Logic

Idea

- Model Decision Based On:
- Total word count of input
- Desired summary length
- Input SizeSummary Size
- Short
- ~12–15%
- Medium
- ~20%
- Long
- ~35–40%
- This ensures:
- ✓ Small text → short summary
- ✓ Large text → proportionate summary



Step 7 – Natural Language Flow

Approach

- Connect idea units using:
- “and”
- “while”
- “as a result”
- Adjust capitalization automatically

Why?

by doing this the model more human like or it summaries will sound very mechanical and robotic



the demo

input

In the city of Veridale, power was measured not by wealth but by influence. Mira, a quiet librarian, discovered a set of forbidden records revealing how the city council manipulated public opinion for decades. As she shared the truth with journalists and activists, she became a target of political threats and social isolation. Despite the risks, Mira continued her efforts, ultimately sparking reforms that reshaped Veridale's governance and restored public trust.



output

In the city of Veridale, power was measured not by wealth but by influence., and despite the risks, Mira continued her efforts, ultimately sparking reforms that reshaped Veridale's governance and restored public trust..

this is done in long setting



the demo

TinyWorld AI – Offline Summarizer

In the city of Veridale, power was measured not by wealth but by influence. Mira, a quiet librarian, discovered a set of forbidden records revealing how the city council manipulated public opinion for decades. As she shared the truth with journalists and activists, she became a target of political threats and social isolation. Despite the risks, Mira continued her efforts, ultimately sparking reforms that reshaped Veridale's governance and restored public trust.

Short

Medium

Long

In the city of Veridale, power was measured not by wealth but by influence., and despite the risks, Mira continued her efforts, ultimately sparking reforms that reshaped Veridale's governance and restored public trust..

Performance & Constraints



- ✓ **Lightweight**
- ✓ **Fast**
- ✓ **Interpretable**
- ✓ **Offline**
- ✓ **Presentation-safe**

⚠ **No deep semantic understanding**

⚠ **No world knowledge**

⚠ **Cannot infer implicit meaning**



Thank you