(Affiliated to University of Mumbai)	***End Semester Examination (R-24/R-25) SH 2026 Answer Key with marking scheme***
Branch: IT	Course: Information Retrieval System
Year/ Semester: III/V	Course code: ITC503
Time: 02 hours	Marks: 60

		Marks
Q. 1	All questions are compulsory.	
A.	Information retrieval system definition 2 Mark Its objectives 3 Marks	05
B.	Explanation of quadruple D, Q, F, R(qi,dj) 5 Marks	05
C.	pattern-matching queries explanation 3 Marks example 2 Marks	05
D.	Stemming explanation 3 Marks Example 2 Marks	05
E.	sequential searching explanation 5 Marks	05
F.	Support for Visual Scannability, Recognition Over Recall, User Control and Freedom, Aesthetic and Minimalist Design, Informative Error Handling and Recovery Loops with explanation 5 Marks	05
Q.2A	Working of vector space model and Boolean model 2 Marks how the vector space model overcomes the exact-matching limitation of the Boolean model , example 3 mark	05
	OR	
Q.2B.	**1. IDF Calculation (2 Marks)** • **1 Mark:** Correctly identifying the document frequency (n_t) for both query terms and stating the correct formulas ($n_{\text{organic}} = 1, n_{\text{soil}} = 2$). • **1 Mark:** Accurately calculating the final log values ($\text{IDF}_{\text{organic}} \approx 0.3010$ and $\text{IDF}_{\text{soil}} = 0.0000$). ◦ *Partial Credit:* Deduct 0.5 marks if a student accidentally uses natural log (ln) or base-2 log ($\log_2$) instead of the requested base-10 log ($\log_{10}$). **2. TF Extraction for Document 1 (1 Mark)** • **1 Mark:** Correctly stating the raw term counts for Document 1 ($\text{TF}_{\text{organic}} = 1, \text{TF}_{\text{soil}} = 1$). **3. TF Extraction for Document 2 (1 Mark)** • **1 Mark:** Correctly identifying the raw term counts for Document 2, especially capturing that "soil" appears twice ($\text{TF}_{\text{organic}} = 0, \text{TF}_{\text{soil}} = 2$). **4. Summary / Final Synthesis (1 Mark)** • **1 Mark:** Consolidating the numbers cleanly (into a matrix, list, or table) showing how the final weights interact, or pointing out the architectural insight that the term "soil" provides zero structural ranking power because its IDF is 0.	05

<table>
<tr><td>Q.3A</td><td>various types of metadata
Preservative, descriptive, technical, structural, administrative with explanation 1 mark each- 5 Marks</td><td>05</td></tr>
<tr><td colspan="3">OR</td></tr>
<tr><td>Q.3B</td><td>
<table>
<tr><th>Media Type</th><th>Data Structure</th><th>Semantic Density</th><th>Temporal Dimension</th><th>Computational Footprint</th></tr>
<tr><td>Text</td><td>Discrete, symbolic tokens (words/characters).</td><td>Extremely High (Words explicitly map directly to concepts).</td><td>Static (Non-linear reading / instant retrieval).</td><td>Very Small (Kilobytes per standard document).</td></tr>
<tr><td>Image</td><td>Continuous 2D pixel grids (spatial arrays).</td><td>Low to Medium (Raw color pixels lack inherent text meaning).</td><td>Static (Spatial coordinates only).</td><td>Medium (Megabytes per item).</td></tr>
<tr><td>Audio</td><td>Continuous 1D amplitude waveforms.</td><td>Low (Audio signals contain sound data, not semantic labels).</td><td>Dynamic (Linear streaming / time-bound).</td><td>Medium to Large (Megabytes to Gigabytes).</td></tr>
<tr><td>Video</td><td>Hierarchical 3D arrays (Spatial grids over time).</td><td>Extremely Low (Highest raw data mass, lowest structural text clues).</td><td>Dynamic (Linear streaming / time-bound).</td><td>Massive (Gigabytes to Terabytes).</td></tr>
</table>
4 Marks
difficulties each create 1 Mark
</td><td>05</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>Q.4A</td><td>(i) components of the college search engine- Localized Web Crawler, The Text Processing & Indexing Engine, The Query Evaluation & Retrieval Engine 3 Marks

Retrieval process- query input, query preprocessing, dictionary lookup, posting retrieval, Statistical Scoring, Interface Rendering 2 Marks

(ii) Evaluate the challenges - The engine must use incremental crawling with HTTP conditional requests, Implement fingerprinting algorithms (like SimHash or MD5 Checksums) at the text-processing stage, The crawler must act as a live link checker and explanation of each 5 Marks</td><td>05

05</td></tr>
<tr><td colspan="3">OR</td></tr>
<tr><td>Q.4B</td><td>(i) Working of search engine 2.5 marks
Working of web browser 2.5 Marks

(ii) Answer: Short-circuit evaluation dictates that in an OR operation V if the first condition evaluates to true, the subsequent conditions are skipped. When the engine encounters the word “protest” it flags that specific clause as valid immediately. 3 Marks

Benefit: It prevents the engine from spending CPU cycles performing string-matching for “boycott” on the rest of the text line, saving processing time 2 Marks</td><td>05

05</td></tr>
</table>

Q.5A	(i) Core Diagnosis & Mechanism Mapping 1 Mark Implicit Feedback via Clicks 1.5 Marks Explicit Feedback via "More Like This" 1.5 Marks Result / Business Impact Synthesis 1 Mark (ii) Interface Component Layout 2 Marks Content Architecture & Visual Layout 1 Mark HCI Design Principles Justification 2 Marks	05 05
	OR	
Q.5B	(i) structural queries to find: (a) sections titled "Stemming", (b) chapters that contain the word "index" in any section, and (c) books with a chapter title containing "web". 3 Marks type of structure each query relies on. 2 Marks (ii) Explanation of why your query specification will structurally reject this match, even though the literal text string is identical. 5 Marks	05 05
	*********** All the Best***********	