

Každá otázka má právě jednu správnou odpověď. Pro každou stránku testu sečtete čísla na levé straně vybrané odpovědi a napíšete součet dole na stránku u Σ . Vydělte součet 7 a zapište celočíselný výsledek. Nezapomeňte, je za to prémiový bod!

1. Google dosahuje při vyhodnocení dotazu rychlosti (vzhledem k množství prohledávaných stránek)
 - lineární
 - logaritmické
 - konstantní

2. LZW algoritmus je
 - semiadaptivní slovníková metoda založená na frázích
 - široce (GIF, PDF) používaná kompresní metoda s rostoucím slovníkem nic z výše uvedeného

3. V rámci předmětu P030 Textové informační systémy jsme se věnovali zejména
 - ektosystému TIS
 - ontosystému TIS
 - endosystému TIS

4. Úrovně složitosti zpracování textů v TIS od nejjednodušší po nejsložitější jsou (sémantická analýza (SA), morfologická analýza (MA), vyhledávání řetězců, kdy slova jsou řetězce písmen (RA), pragmatická analýza (PA)):
 - MA/SA/RA/PA.
 - RA/MA/SA/PA.
 - RA/MA/PA/SA.
 - PA/RA/MA/SA.

5. Mezi modely dokumentů TIS nepatří
 - Pravděpodobnostní model.
 - Model shluků dokumentů.
 - Kontextový model.

6. Dvouúrovňové vrstvené kódování je

- signaturová metoda.
prefixové kódování založené na vrstvení frází.
metoda dekompozice zápisu slovníku při LZC kompresi.

7. Který z následujících výroků je nepravdivý?

Koeficient podivnosti slova je míra neobvyčejnosti slova v daném jazyce, používaná při kontrole textu.

Kompresa založená na kontextovém modelování je náročná na paměť a čas.

- Huffmanovo kódování je zobecněním Shannon-Fano kódování.
Každý regulární jazyk lze reprezentovat regulárním výrazem.

8. Které z následujících tvrzení je pravdivé?

Entropie kódové zprávy nad jednoprvkovou abecedou je nenulová.

Aritmetické kódování je afixový kód.

Každý Shannon-Fano kódový strom má sourozeneckou vlastnost.

- Eliasův kód $\omega(N)$ může být i delší než Fibonacciho kód pro totéž N .

9. Které z následujících tvrzení je pravdivé?

- Google si drží lokálně uložené kopie textů indexovaných dokumentů celého Internetu.

Paiceův model je jedna z modifikací vektorového modelu dokumentů.

Rychlost vyhledávání v desetimilionovém lexikonu slov implementovaného pomocí trie na Pentiu nedosahuje 100000 nalezených slov za vteřinu.

10. Které z následujících tvrzení je pravdivé?

- Analytické kódování znamená zaznamenání posloupnosti rozhodnutí provedených syntaktickým analyzátořem.

Protisměrné vyhledávání je vždy rychlejší než sousměrné.

Unixový program compress používá Huffmanovo kódování.