

長崎大学工学部情報システム工学科
平成 19 年度 AO 入試課題
課題 2

この課題では文章ファイルから文字を検索するときに使う命令記述形式の一つである**正規表現**というものを使って、ある文章から目的の語句を探してもらいます。

まず、正規表現を説明した**添付資料**に目を通して下さい。目の前の計算機にはこの課題のための正規表現を実行する**プログラム**が用意されており、あなたの入力を待っているページが表示されています。このプログラムでは検索対象として、解答に用いる**本番データ**と、実行結果がいつでも画面に表示されるので考え方を確かめることができる**実験データ**の二つが用意されています。どちらのデータも自由に検索ができますので、プログラムを使いながら添付資料を読み、以下の設問に答えて下さい。

「**練習問題**」と題された問題は理解を助けるための問題です。解答する必要はありません。

練習問題 1 「Holmes」という文字列を含む行は実験データで 1 行、本番データで 457 行あります。確かめなさい。

練習問題 2 「Holmes」という文字列を含む行は実験データで 1 行、本番データで 423 行あります (先の問題とは先頭に空白文字があるところが違います)。確かめなさい。

問 1 「Holmes」という文字列で始まる行は本番データで何行ありますか。

問 2 練習問題 2 の解答と問 1 の解答を足すと (全ての場合に一致するはずの) 練習問題 1 の解答になりますか。ならないなら、どうして差が生じたと思われるか説明して下さい。

問 3 「the」という文字列が少なくとも 2 回出てくる行は実験データには 2 行ありますが、本番データには何行ありますか

問 4 本番データには「*」という文字は出てきますか。出てくるなら実例を 1 行選び、そのまま解答欄に写してください。

問 5 本番データの中で、「Sherlock」という文字列が出てくる行の中で「Holmes」という文字列が出てこない行は何行ありますか。どのように求めたか、使った解き方も記入して下さい。
ヒント: 本番データの中にはこのような行があります。

"Good-morning, madam," said Holmes cheerily. "My name is Sherlock Holmes. This is my intimate friend and associate,

練習問題 3 z という文字の直後に a から z の中の任意の英文字が一つ以上続くパターンが出てくる行は実験データには 0 行、本番データ中には 113 行あります。確かめなさい。

参考: 「zi」, 「zaaa」, 「zab」, 「abczef」 はどれも題意に合います。

問 6 本番データ中で、z という文字で始まる一番長い単語はなんですか。なお、単語は英文字 (a から z および A から Z) のみから構成され、単語と単語とはスペースまたは行が変わることによって区切られます。

問 7 本番データの中で、「Dr.」で始まる人の名前を全て挙げなさい。

長崎大学工学部情報システム工学科
平成 19 年度 AO 入試課題
課題 2

問 8 本番データの中で、「Holmes」と「Watson」という文字列が共に含まれる行は何行ありますか。

問 9 h という文字の後に a から z の中のどれかの文字が続き、さらに a から z の中のどれかの文字が続くというパターンにマッチする同一文字列が 2 回出てくる行は実験データには 2 行ありますが、本番データには何行ありますか。

例えば,

aaa chat bbb shat

では h の後に a から z の中のどれかの文字が続きさらに a から z の中のどれかの文字が続く「hat」が 2 回出てくるので題意に沿った行になりますが,

aaa ahaz bbb bhab

では「haz」と「hab」はそれぞれパターンにはマッチしていますが、この行は同一文字列が 2 回出てくるという条件を満たしていないので題意には沿いません。

練習問題 4 h という文字で始まる単語でその h の後に a から z の中のどれかの文字が続きさらに a から z の中のどれかの文字が続くというパターンにマッチする同一文字列が少なくとももう一回出てくる行という条件に、実験データ中では以下の 2 行が該当します。確かめなさい。

To Sherlock Holmes she is always THE woman. I have seldom heard him mention her under any other name. In his eyes she a doubt upon all his mental results. Grit in a sensitive instrument, or a crack in one of his own high-power lenses,

問 10 h という文字で始まる単語でその h の後に a から z の中のどれかの文字が続きさらに a から z の中のどれかの文字が続くというパターンにマッチする同一の単語が少なくとももう一回出てくるという行は実験データ中に 1 行ありますが、本番データ中には何行ありますか。(前問と違い、2 回目の出現は別の長い単語の中に含まれてはいけません。)

問 11 本番データ中で、1 行に出てくる回数をもっとも多い単語はなんですか。その回数も教えてください。

問 12 本番データ中で、同じ文字が 4 回以上出てくる単語を一つ見つけなさい。