

# Varianta A

## *Sumarizator de fișiere Markdown*

### 1 Descrierea problemei

Se dă un fișier Markdown<sup>1</sup>. Se cere realizarea unei sumarizări simpliste a textului din fișier, după regulile prezentate mai jos.

**Atenție!** Există numeroase arome (*flavors*) de markdown. Vom utiliza în mod **exclusiv** specificațiile din specificația originală de markdown<sup>2</sup>, chiar dacă nu vom utiliza *toate* aceste specificații.

### 2 Reguli pentru sumarizare

Orice element din output va fi indentat cu un număr  $3 * n$  de spații, unde  $n$  este cel puțin 1. Vom numi  $n$  nivel de indentare.

Un **titlu** (*header*) de nivel 1 sau 2 va fi indentat la nivel 1, respectiv 2. Toate elementele subordonate titlului respectiv vor fi indentate cu cel puțin același nivel de indentare. Un titlu cu textul “Paragrafe” va fi prezentat la ieșire ca “+ Paragrafe”. Vor fi suportate atât titlurile în format *atx* cât și *Setex* (vezi referința de sintaxă).

Titlurile în format *Setex* vor fi considerate ca atare numai dacă sublinierea este de *aceeași lungime* ca și conținutul titlului. Pentru titlurile în format *atx* “închise” (care au un număr oarecare de # la sfârșit) nu se va afișa și închiderea.

Un **titlu de nivel 3** sau mai mare va fi indentat la nivelul 3, și va fi afișat la fel ca un titlu de nivel 1 sau 2.

Un **paragraf** va fi reprezentat doar prin prima propoziție din paragraf (până la un semn de punctuație diferit de virgulă sau până la o linie goală), încadrată de :. Afișarea sumarizării paragrafului în ieșire se va plasa la nivelul de indentare curent. Un paragraf cu textul “I am hungry. I will eat some chicken. With fries.” va fi afișat în ieșire ca :I am hungry:. Sumarizarea unui paragrafului va conține, în afară de prima propoziție, și orice alte elemente accentuate sau linkuri. Astfel, un paragraf cu textul “I am \*so hungry\* today. I will eat some \*\*chicken\*\*. With \_fries\_.” va fi afișat în ieșire ca :I am <so hungry> today <chicken> <fries>:. Observați menținerea marcării elementelor accentuate și în prima propoziție.

---

<sup>1</sup><https://en.wikipedia.org/wiki/Markdown>

<sup>2</sup>Markdown Syntax Documentation <https://daringfireball.net/projects/markdown/syntax>

Orice element de **accentuare** a textului (folosind “\*, \*\*, -, --”) va fi tradus în ieșire ca textul respectiv între paranteze unghiulare (vezi mai sus).

Un **link** de forma [an example](http://example.com/ "Title") va fi afișat în sumarizare ca [@an example]. O **image** de forma ![Alt text](/path/to/img.jpg "Optional title") va fi sumarizată ca [!Alt text].

**Listele** (numerotate sau nu) vor fi sumarizate ca o enumerare a elementelor din listă, încadrată de paranteze drepte, fiecare element fiind sumarizat prin primele maxim 3 cuvinte din element. Elementele se despart prin ;. Pentru fiecare element, se adaugă și orice text accentuat sau linkuri care urmează în contextul aceluiași element, la fel ca și la paragrafe. De exemplu, lista

```
* First, get these ingredients:
* Boil some water.
* Don't get burned because the water *is hot*

    Always remember to be careful when the water is _hot_.
* Dump
  everything
  in the pot and follow
  this algorithm:
```

va fi sumarizată ca [First, get these; Boil some water; Don't get burned <is hot> <hot>; Dump everything in]

Sumarizarea unei liste va fi afișată la un nivel de indentare incrementat cu 1 față de nivelul curent. Dacă elementul din listă are mai puțin de 3 cuvinte, dar mai are alte paragrafe, sumarizarea nu va include și cuvinte din celelalte paragrafe.

**Citatele** (*blockquotes*) vor fi sumarizate ca primele maxim 5 cuvinte din citat, încadrate de ". Sumarizarea unui citat va fi afișat la un nivel de indentare incrementat cu 1 față de nivelul curent. Dacă citatul apare în interiorul unei liste, sumarizarea lui va fi adăugată în sumarizarea elementului de listă în care apare.

Orice **bloc de cod** va fi sumarizat ca [code]. Dacă blocul apare într-o listă, sumarizarea lui va fi adăugată în sumarizarea elementului de listă în care apare; altfel, sumarizarea va fi adăugată pe o linie separată, la nivelul de indentare curent.

Dacă apar elemente de cod în paragrafe, liste, etc, este la alegere dacă dacă îl ignorați, îl afișați, sau nu mai procesați cuvintele imediat următoare.

O linie de separare (*hr*) va fi reprezentată ca o serie de slash-uri: //////////////

### 3 Detalii de implementare

Ieșirea se va afișa pe măsură ce se parcurge fișierul; nu se va reține niciun element din input pentru afișare ulterioară, în afară de titlurile de tip *atx*.

Puteți folosi C pentru a gestiona nivelul de indentare din afișare.

## 4 Exemple

Intrare 1:

```
# main title

## title

### other title
word1 word2 word3 word4 **hello** word 5 _hello_

really this should be a text with meaning.

The meaning is in the words
on one or two
or serveral lines in the paragraph. There are more sentences in a paragraph.

Or use a _graph_.

Here is a short list:

* one item
* two items
* three items

here1
=====

here2
-----

-----

-----

bla

> there is
> more meaning
> in an image than
> in 1000 words

the end.
```

Ieșire 1:

```
+ main title
+ title
+ other title
:word1 word2 word3 word4 <hello> word 5 <hello>:
:really this should be a text with meaning:
:The meaning is in the words on one or two or serveral lines in the paragraph:
:Or use a <graph>:
:Here is a short list:
[one item; two items; three items]
+ here1
+ here2
//////////
//////////
:bla:
"there is more meaning in"
:the end:
```

*Alte exemple se găsesc în arhiva cu teste.*