

Task	Subtask	Subtask description	Person	Hours	Price
<p>A Divány-korpusz egy részének (3000 mondatnyi szöveg) annotálása, és a nyertes pályázó által megírt kombinációs szabályok hatékonyságának a tesztelése ezen a korpuszon. Cél: Egy olyan korpusz létrehozása, amelyben a mondatok szintjén meg adva a szentimentérték, és ennek a segítségével a különböző elemzési megoldások eredményességét mérni lehet.</p>	1. minőségellenőrzés	A korpusz annotálni kívánt részének a végigellenőrzése, az annotálási hibák javítása	annotátorok	20 óra	
	2. az egyetértésméréshez pilot-annotálás	A két annotátor annotálja ugyanazt a korpuszrészletet, hogy az egyetértést mérni tudjuk	annotátorok	20 óra	
	3. egyetértésmérés	Az annotáláson a statisztikus egyetértést mér Kappa-statisztikával	statisztikus	5 óra	
	4. annotálás	A korpusz annotálása	annotátorok	50 óra	
	5. a létrehozott korpusz elemzése a szótárakkal, és a hatékonyság mérése	A kézzel annotált korpusz elemzése a szentimentszótárral, és a szótáras elemzés eredményének az összevetése a kézi annotálással	statisztikus	8 óra	
	6. A kombinációs szabályok megírása + elegendő tesztet összegyűjtése a korpuszból	BNF grammar formátumban fel kell írni a szabályokat, hogy a fejlesztő könnyen és gyorsan tudjon haladni. Teszt esetek itt: https://docs.google.com/spreadsheets/d/10Sm0LcrbEEhgSadshJ9Cv73dD8zfd1HqUAon1j2HK0k/edit#gid=0	szakértő	20 óra	
	7. A nyertes pályázó által megírt kombinációs szabályok kódolása		fejlesztő	65 óra	85 óra

	<p>8. A korpusz elemzése a szótárak és a szabályok kombinációjával, majd a hatékonyság mérése az egyszerű szótáras módszerhez képest</p>	<p>A kézzel annotált korpusz elemzése a szótárakkal és a kódolt szabályokkal együtt, majd az eredmény összevetése a kézi annotációval és az egyszerű szótáras elemzés eredményével</p>	<p>statisztikus, fejlesztő</p>	<p>12+12= 24 óra</p>	
<p>Szentimentelemző demó létrehozása (szótár- és szabályalapú). Cél: egy olyan program létrehozása, amelyen meg lehet nézni, hogy a szótárainkkal és a megírt szabályainkkal hogyan elemzünk egy tetszőlegesen beírt mondatot szentimentre (score-adás)</p>	<p>Fejlesztés</p>	<p>Fejlesztő: Eszközben Pyparsing vagy ANTLR kell.</p>	<p>fejlesztő</p>	<p>35 óra</p>	<p>50 óra</p>
<p>Szóasszociációs adatbázis létrehozása (ehhez szükséges: fejlesztő a tervezéshez, reklám az adatgyűjtéshez, fejlesztő az adatok feldolgozásához). Cél: Egy olyan szóasszociációs adatbázis létrehozása, amely a szentimentelemzésben problémás elemeket, és az ezekre adott asszociációkat tartalmazza (értékváltás és értékvesztés).</p>	<p>1. Tervezés 2. A kérdőív összerakása 3. Megfelelő mennyiségű adat összegyűjtése 4. Az adatok kiértékelése</p>	<p>(mik a seed adatok, hogyan ellenőrizzük a bekerülő új adatokat, hogyan dobjuk ki a felhasználónak a kérdéseket stb.) Egy egyszerű oldalra lenne szükség, amely segítségével adatokat gyűjthetünk (pl. egy Flask, PHP keretrendszer).</p>	<p>statisztikus fejlesztő</p>	<p>40 óra 55 óra</p>	<p>- -</p>