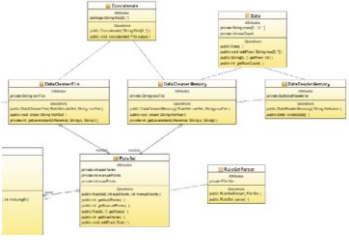
	<p align="center"><b>Adattisztítási algoritmusok kutatása és kapcsolódó módszertan fejlesztése</b></p>
<p><b>Projektvezető:</b></p> <ul style="list-style-type: none"> <li>• Martinek Péter</li> </ul>	<p><b>Cél:</b> Az ipari partner által rendelkezésre bocsátott ügyféladat állományok vizsgálata adatminőségi szempontból. További cél adattisztítási megközelítés kidolgozása, algoritmusok és módszertan tervezése és végrehajtásának előkészítése.</p>
<p><b>Ipari megbízó:</b></p> <ul style="list-style-type: none"> <li>• Magyar Posta Biztosító Zrt.</li> </ul>	<p><b>Eredmények:</b> A kutatás során a rendelkezésre álló adatállományokat tanulmányoztuk. Megállapítottuk, hogy az adatok minősége időszakonként és forrásrendszerrel függően változó. Az előforduló hibákat osztályoztuk, azokra külön detektálási algoritmusokat s javítási módszertant javasoltunk. A kutatás központi témája a duplikátumok kiszűrése és megszüntetése volt, melyre félautomata megoldást terveztünk és fejlesztettünk.</p>
<p><b>Időtartam:</b> 2010.06.28.-2010.09.15.</p>	

	<p align="center"><b>Researching data clearing algorithms and development of a data clearing method</b></p>
<p><b>Project Leader:</b></p> <ul style="list-style-type: none"> <li>• Martinek Péter</li> </ul>	<p><b>Aim:</b> Analysing the customer data provided by the sponsor to estimate general data quality. Furthermore the project proposed an approach to detect data quality failures and a complex method containing algorithms to eliminate them.</p>
<p><b>Industrial partner:</b></p> <ul style="list-style-type: none"> <li>• Magyar Posta Biztosító Zrt.</li> </ul>	<p><b>Results:</b> During the research the provided data have been deeply analysed. IT has been stated that the quality of the stored data is rather varying depending on the date of storing and the type of its source application system. The failures were classified and different solutions were provided for the detection and the elimination for each class. The main focus was set to the detection and elimination of duplicated entries where a half-automated solution was provided as the main result of the researches.</p>
<p><b>Duration:</b> 2010.06.28.-2010.09.15.</p>	