



## Roadmap for natural language processing (NLP)

Description and state of the art	
 Definition	<p>Subfield of artificial intelligence and based on the advances in linguistics and machine learning, and closely related with Bots. <i>Natural Language Processing (NLP)</i> is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human-computer interaction[189]. NLP technology involves the ability to turn text or audio speech into encoded, structured information, based on an appropriate ontology[190].</p> <p>NLP solutions enable communication between human and machine by analysing the content written and spoken in natural human language and converting it into the machine understandable language[191]. Individual challenges within NLP involve natural language understanding, enabling computers derive meaning from human or natural language input, natural language generation, etc.</p>
 Addressed societal /business or public sector need	<p>Public sector need:</p> <p>Digitization</p>
 Existing solutions /applications /services	<p>There are several NLP solutions on the market, such as:</p> <ul style="list-style-type: none"> <li>• Clarabridge NLP[192]</li> <li>• RASA NLU[193]</li> <li>• Ignitho NLP[194]</li> <li>• Innoetics[195]</li> <li>• NLP Technologies[196]</li> <li>• Data Genic NLP[197]</li> <li>• Vocali[198]</li> </ul>
 Main actors regarding R&D	<ul style="list-style-type: none"> <li>• University Of Edinburgh</li> <li>• Centre National de la Recherche Scientifique, University of Oxford</li> <li>• University of Cambridge</li> <li>• Athena Research and Innovation Center in Information Communication &amp; Knowledge Technologies</li> </ul>

of this technology	<ul style="list-style-type: none"> <li>• Stichting Katholieke Universiteit</li> <li>• The University of Sheffield</li> </ul>
 Current research activities	There are overall 151 projects related to 'natural language processing'. Projects which set the focus on 'natural language processing' itself are <ul style="list-style-type: none"> <li>• <b>HeLeNLP</b> ("Heterogeneous Learning for Natural Language Processing"), with the goal to design new paradigms for large-scale learning of natural language problems in various languages from heterogeneous data sources of variable size, quality, amount of supervision and type[170].</li> <li>• <b>JointStructuredPred</b> (Machine Learning Methods for Complex Outputs and Their Application to Natural Language Processing and Computational Biology)[199].</li> <li>• <b>GRAMPLUS</b> (Grammar-based Robust Natural Language Processing), to restore grammatical theory to its necessary place in the theory of human language behaviour, by providing a more restricted theory of constructions than others on offer[200].</li> </ul>
 Impact assessment	<b>Public sector modernization:</b> <ul style="list-style-type: none"> <li>• Sustainability</li> <li>• Quality of Services Provided</li> <li>• Level of Participation</li> </ul> <b>Public Sector as an Innovation Driver:</b> <ul style="list-style-type: none"> <li>• Social equity and inclusiveness</li> <li>• Privacy &amp; Security</li> </ul>
<b>Necessary technological modifications</b>	
 Potential cases use	Potential applications of the NLP technology which could cater for the need of the public sector for further digitization include: <ul style="list-style-type: none"> <li>• Conversational interfaces</li> <li>• Automated online assistants</li> <li>• Sentiment analysis</li> <li>• Native language identification</li> <li>• Internet Bots (Robots)</li> </ul>
 Technological challenges	Challenges in natural language processing frequently involve natural language understanding, natural language generation (frequently from formal, machine-readable logical forms), connecting language and machine perception, dialog systems, or some combination thereof[189]. <p>Natural language processing is successful in meeting the challenges as far as syntax is concerned. But it still has to go a long way in the areas of semantics and pragmatics. The issues still unresolved in semantics are finding the meaning of a word or a word sense, determining scopes of quantifiers, finding referents of anaphora, relation of modifiers to nouns and identifying meaning of tenses to temporal objects[201].</p>

	<p>Beside the aforementioned theoretical challenges, in NLP, there is also a huge gap between the availability of the solution in general and the availability of a reliable product implementing the solution well enough for the real world, and thus being effective for more than one specific domains, languages, means of expression (indicatively social media, etc.) and both short messages and very long sentences.</p>
<p><b>Necessary activities (in or for the public sector)</b></p>	
 <p>Development of a specific training necessary</p>	<p><b>Open task</b></p> <p>End user expertise and thereby training is required in corpus analysis or computational linguistics and data science so as to effectively analyse text/speech and build efficient models and ontologies.</p>
 <p>Advanced adapted infrastructure needed or ICT</p>	<p><b>Open task</b></p> <p>The development of domain-specific ontologies or language specific dictionaries is required.</p>
 <p>Change of (public sector internal) processes necessary</p>	<p></p> <p>No change of public sector internal processes is necessary.</p>
 <p>Promotion information / of stakeholders necessary</p>	<p></p> <p>No promotion / information provision to stakeholders is necessary.</p>
 <p>Need to deal with cyber security issues</p>	<p></p> <p>No cyber security issues identified.</p>

 <p>New or modified legislative framework or regulations necessary</p>		<p>No modifications in the legal framework are required.</p>
 <p>Development of a common standard necessary</p>		<p>No standards' development is necessary.</p>
 <p>Need for a more economical solution</p>	<p><b>Open task</b></p>	<p>There may be the need for more economical solutions in the sense that advanced infrastructure is as well needed, accompanied with big efforts in training systems on recognising language and speech.</p>
<p><b>Dealing with challenges</b></p>		
 <p>Ethical issues</p>		<p>No ethical issues identified.</p>
 <p>Societal issues</p>		<p>No societal issues identified.</p>
 <p>Health issues</p>		<p>No health issues identified.</p>
 <p>Public acceptance</p>		<p>No public acceptance issues identified.</p>