

Deutsche Version (siehe Unten)
Version française (ci-dessous)

Lay Summary

Project title	NLP-powered mapping of clinical reports onto SNOMED-CT concepts for tumour classification (NLPforTC)
Main applicant	Dr. Thomas Fabbro, Department Clinical Research, University Hospital Basel.
Consortium	University Hospital Basel, University Hospital Zurich.
Short Summary	Routine radiology and pathology reports are a valuable source of information for clinical research project. Within this infrastructure development project, we will implement and evaluate natural language processing (NLP) based methods in order to extract standardized information from unstructured clinical documents.
Background	Using information from clinical reports for research purposes often requires a retrospective, manual coding and structuring of the data. This process is time-consuming and mostly tailored to the respective research project, hence it is not easily applicable to other purposes. Numerous studies, especially from the Anglo-Saxon area, have shown that such tasks can be efficiently solved by natural language processing (NLP) and machine learning techniques. However, such procedures need to be adapted to the medical specialty, the local language and specific reporting guidelines. We propose to evaluate and apply NLP, machine and statistical learning methods in order to extract standardized information from radiology and pathology reports of cancer patients related to tumour classification.
Goal	The primary goal of this infrastructure development project is to implement and evaluate a NLP-driven process to map information from radiology and pathology reports related to tumour onto the SNOMED-CT terminology. Based on the extracted information, a Tumour-Node-Metastasis (TNM) classifier will be developed in order to predict the tumour class of individual patients.
Significance	The proposed project addresses a main challenge to further develop personalized health research, namely provide standardized clinical information gathered from health care documents. Currently this information is not standardized and often stored as unstructured text. The application of SNOMED-CT in the context of TNM classification could lead to a more frequent usage of this terminology within routine clinical documentation and hence enhance the usability of clinical information for research in general. Since a manual mapping of SNOMED-CT concepts is very time-consuming, an automatic NLP-driven procedure would have a great impact on processing and dissemination such information for research.

D

Deutsch

Projekttitlel	NLP-basierte Methoden zur Bestimmung von SNOMED-CT Konzepten und zur Tumorklassifikation (NLPforTC)
Hauptgesuchssteller	Dr. Thomas Fabbro, Departement Klinische Forschung, Universitätsspital Basel.
Consortium	Universitätsspital Basel, Universitätsspital Zürich.
Kurzzusammenfassung	Klinische Verlaufs- und Befundberichte enthalten wertvolle Informationen die auch für Forschungszwecke genutzt werden können. Im Rahmen dieses Infrastruktur Entwicklungsprojektes möchten wir computer-linguistische Verfahren nutzen, um standardisierte Informationen aus unstrukturierten Textdokumenten zu extrahieren.
Hintergrund	Um spezifische Informationen aus klinischen Berichten herauszufiltern, müssen diese üblicherweise von Mediznern gelesen, strukturiert und interpretiert werden. Diese manuelle Vorgehensweise ist zeitaufwändig und die dabei gewonnen Informationen oft nicht wiederverwendbar für andere Forschungszwecke. Eine Reihe von Studien, vor allem im englischsprachigen Raum zeigen, dass mit Methoden aus dem Bereich des Natural Language Processing (NLP) Informationen aus Textdokumenten effizient und zuverlässig extrahieren werden können. Entsprechende Algorithmen müssen jedoch für den jeweiligen Sprachraum und das medizinische Fachgebiet angepasst werden. Die Entwicklung und Etablierung von NLP-basierten Methoden ist daher von grosser Bedeutung, um effizient Informationen aus klinischer Dokumentation zu gewinnen.
Das Ziel	Ziel dieses Infrastruktur Entwicklungsprojektes ist es, NLP-basierte Prozeduren zu entwickeln, um Tumorinformation die in Radiologie- und Pathologieberichten enthalten sind, in eine kodierte und strukturierte Form gemäss SNOMED-CT Terminologie zu übersetzen. Basierend auf diesen Informationen wird anschliessend die Tumor Klasse jedes Patienten automatisch abgeschätzt.
Bedeutung	Dieses Projekt adressiert eine bedeutende Herausforderung im Rahmen der Informationsgewinnung aus klinischer Routine-Dokumentation: Verlaufs- und Befundberichte werden oft als Freitext und damit nicht ausreichend strukturiert erfasst um maschinell weiterverarbeitet werden zu können. Die Übersetzung auf SNOMED-CT Konzepte ermöglicht hingegen eine standardisierte und vielseitige Auswertung dieser Informationen, beispielsweise für Forschungs- und Qualitätsmanagementfragen. Da die manuelle Zuweisung von SNOMED-CT Konzepten sehr zeitaufwendig ist, sind NLP-basierte Methoden für eine praktische Anwendung im klinischen Alltag erforderlich.

Français

Titre du projet	Cartographie des rapports cliniques alimenté par la NLP sur les concepts SNOMED-CT pour la classification des tumeurs (NLP-powered mapping of clinical reports onto SNOMED-CT concepts for tumour classification ; NLPforTC)
Requérant principal	Dr. Thomas Fabbro, Département de Recherche Clinique, Hôpital Universitaire de Bâle.
Consortium	Hôpital Universitaire de Bâle, Hôpital Universitaire de Zurich.
Résumé	Les rapports de radiologie et de pathologie de routine constituent une source précieuse d'informations pour un projet de recherche clinique. Dans le cadre de ce projet de développement d'infrastructure, nous mettrons en œuvre et évaluerons des méthodes basées sur le traitement du langage naturel (NLP) afin d'extraire des informations standardisées à partir de documents cliniques non structurés.
Context	L'utilisation d'informations provenant de rapports cliniques à des fins de recherche nécessite souvent un codage rétrospectif, manuel et une structuration des données. Ce processus prend beaucoup de temps et est principalement adapté au projet de recherche concerné, il n'est donc pas facilement applicable à d'autres fins. De nombreuses études, en particulier dans la région anglo-saxonne, ont montré que de telles tâches peuvent être efficacement résolues par des techniques de traitement du langage naturel (NLP) et d'apprentissage automatique. Cependant, de telles procédures doivent être adaptées à la spécialité médicale, à la langue locale et à des directives spécifiques en matière de signalement. Nous proposons d'évaluer et d'appliquer des méthodes d'apprentissage statistique et automatique de la NLP afin d'extraire des informations standardisées des rapports de radiologie et de pathologie des patients cancéreux liés à la classification des tumeurs.
But	L'objectif principal de ce projet de développement d'infrastructure est de mettre en œuvre et d'évaluer un processus piloté par la NLP pour cartographier les informations provenant des rapports de radiologie et de pathologie liés à la tumeur selon la terminologie SNOMED-CT. Sur la base de l'information extraite, un classificateur Tumor-Node-Metastasis (TNM) sera développé afin de prédire la classe tumorale de patients individuels.
Importance	Le projet proposé aborde un défi majeur pour développer davantage la recherche en santé personnalisée, à savoir fournir des informations cliniques standardisées provenant de documents de soins de santé. Actuellement, cette information n'est pas normalisée et souvent stockée sous forme de texte non structuré. L'application de SNOMED-CT dans le contexte de la classification TNM pourrait conduire à une utilisation plus fréquente de cette terminologie dans la documentation clinique de routine et ainsi améliorer la facilité d'utilisation de l'information clinique pour la recherche en général. Comme une cartographie manuelle des concepts SNOMED-CT prend beaucoup de temps, une procédure automatique basée sur la NLP aurait un grand impact sur le

	traitement et la diffusion de ces informations pour la recherche.
--	---