

## خلاصة رسالة الدكتوراه

The last decade has seen unprecedented growth in both the production of biomedical data and amount of published literature discussing it. Tissue engineering laboratories at *Alexandria University* aim to regenerate dental tissues by tissue engineering principles and technology (dentine formation process). Dentine formation is governed by biological mediators or growth factors (protein) and interactions amongst different proteins. Dentine formation needs the support of continuous updated information about protein-protein interactions. Thus, having a scalable, robust system for protein interaction discovery provides a major information extraction tool for molecular biologists to automatically extract and transfer updated biological data about protein-protein interactions from unstructured form, to a structured form to be used in their respective applications.

Thus in this thesis, we present PIELG: a system for extracting information about protein – protein interactions from abstracts of biomedical papers. The data obtained from this system will be first confirmed partially in their laboratory. Then they will use those extracted information about protein – protein interactions in Dentine formation process. Our approach is based on first splitting abstracts into simple sentences. Then, the system tags biological entities with the help of biomedical and linguistic ontologies. Finally, the system extracts complete interactions by analyzing the matching contents of syntactic roles and their linguistically significant combinations.

PIELG handles complex sentences and extracts multiple and nested interactions specified in a sentence. The scope of our experiments is limited to abstracts describing human protein function. The corpus of the PIELG is selected in order to evaluate the proposed protein-protein interaction validation method. This corpus is selected to be about proteins currently considered to have roles in the *dentine formation* process and involved in dentinogenesis. We performed experimental evaluations of the PIELG systems.

**The interactions extracted by the PIELG system are manually examined for precision and recall. The sensitivity of the system is given by the recall measure, calculated as the ratio between the interactions extracted correctly and the interactions present in text. Precision is a measure of correctness of the system by measuring the number of times the results are extracted correctly in comparison with the total number of results. Our experimental results show that the PIELG system presented here achieves better performance without the need of manual pattern creation (by user) which is required for the other systems.**