inglês na interface do software por serem nomes padronizados.

4. Resultados Obtidos

A amostra de estudantes alvo do estudo foi constituída durante uma atividade realizada na disciplina de Química Orgânica do curso de Engenharia de Materiais e Tecnólogo em Polímeros. A atividade teve o objetivo de expor as ferramentas de visualização à análise de um professor que não fosse da área da informática. Esta atividade foi iniciada com a solicitação aos estudantes da leitura do artigo Polietileno: Principais Tipos, Propriedades e Aplicações (COUTINHO, MELLO e MARIA, 2003). O artigo descreve as principais características e aplicações do polietileno de baixa densidade, de alta densidade, linear de baixa densidade, ultra-alto peso molecular e de ultrabaixa densidade, correlacionando a sua estrutura química com suas propriedades. A ferramenta APT foi aplicada nos textos de sete estudantes com relação ao estabelecimento da relação entre estrutura e propriedade dos vários tipos de polietileno que justifiquem suas aplicações. As produções textuais dos estudantes foram separadas para facilitar as análises pelo tipo de polietileno, gerando uma produção textual para cada tipo de polietileno (cinco tipos) e para cada estudante (sete estudantes). As visualizações geradas foram disponibilizadas ao professor, sendo brevemente descritas a seguir. Na tela do computador as imagens são clicáveis e se modificam conforme a interação do professor avaliador. Cabe destacar que as imagens são exibidas em tamanho grande para o professor e a resolução na tela é consideravelmente melhor do que na versão que consta neste artigo.

A primeira visualização é a Collapsible Tree (Figura 3). As ramificações iniciaram com o tema polímero e tipos de polímeros e após os estudantes da turma que realizaram a atividade. Clicando em cada estudante têm-se uma listagem das palavras mais relevantes utilizadas na sua produção textual. No exemplo selecionou-se “polietileno de alta densidade” e “estudante-poli-3” (padrão de identificador utilizado para os estudantes). Ao descrever as propriedades do polietileno de alta densidade o estudante utilizou termos importantes como: maior, molecular, ramificações, rigidez, tipo, usado. O professor avaliou que o texto do aluno apresentava palavras-chave relevantes, e que a própria visualização contribuiu para avaliar quais as propriedades que o estudante compreendeu e soube inserir em seu texto.

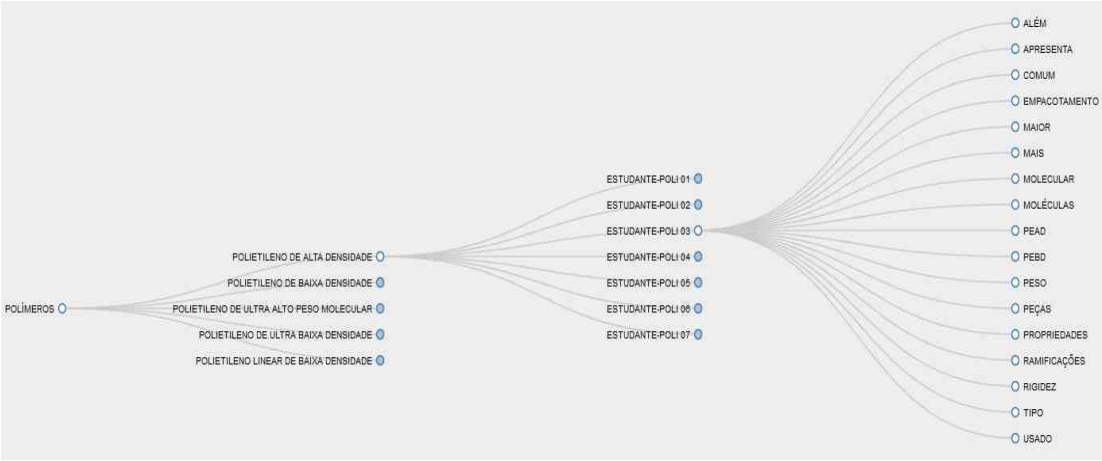

Figura 3 - Collapsible Tree (Árvore Flexível) dos estudantes em todas as atividades.

A Figura 4 apresenta o Bubble Chart com a análise de todas as atividades de um estudante. Nela os macrotemas são representados pela área da bolha e o texto do centro da bolha identifica o macrotema. Os polímeros são caracterizados por cores, onde cada cor representa um tipo de polímero. Ao posicionar o cursor sobre uma das bolhas é

apresentada a palavra completa acompanhada de sua pontuação. A visualização distribui os polímeros com maior volume de termos em macrotemas na região central e na periferia estão os com menos termos. Nesta visualização é possível perceber a representatividade dos termos mais significativos comparando o raio das bolhas. Enquanto o gráfico da Figura 3 facilita a visualização das tarefas de cada aluno individualmente, o gráfico de bolhas possibilita que se visualize todos os alunos em uma tarefa ou todas as tarefas de um aluno. Já a visualização de todas as tarefas de todos os alunos se torna mais complexa. O gráfico de bolhas foi avaliado pelo professor como um bom recurso para avaliar a progressão do estudante em termos do vocabulário empregado, bem como perceber se as propriedades esperadas constam do texto.

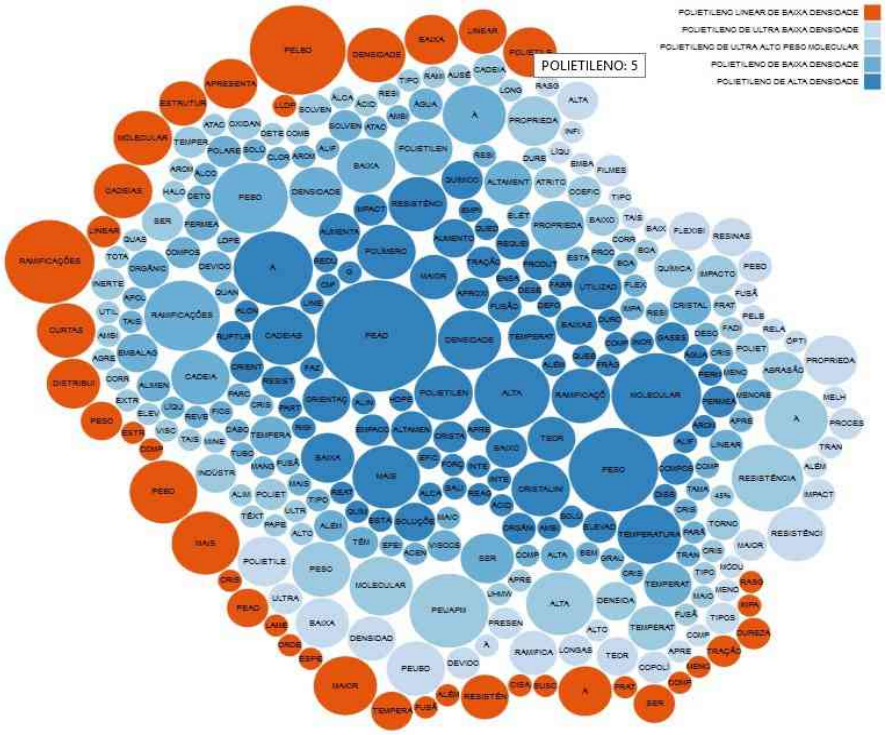

Figura 4 - Bubble Chart (Gráfico de Bolhas) de um estudante em todas as atividades.

A Figura 5 apresenta a visualização Word Cloud. A nuvem de palavras representa a visualização de dados de texto e é representada por palavras isoladas. A importância de cada palavra é mostrada com um tamanho da fonte ou cor. Neste formato pode-se perceber os termos mais proeminentes, sua importância relativa e a frequência da ocorrência das palavras no texto fornecido. Quanto maior for o número de ocorrências de palavras, maior será a nuvem de palavras. Esta visualização não mostra a pontuação, apenas apresenta visualmente, de forma intuitiva, quais as palavras mais relevantes na produção textual do estudante. Observou-se que as palavras mais importantes e de maior frequência com relação à análise do texto na atividade foram: PEAD, peso, molecular e alta.

A Figura 6 ilustra uma das visualizações utilizando Rotating Cluster. Ela é similar a visualização Collapsible Tree, apresentando contudo as palavras ficam em uma disposição diferente. O seu objetivo é gerar a listagem das palavras, não sendo possível saber quais delas são as mais relevantes na produção textual do estudante.

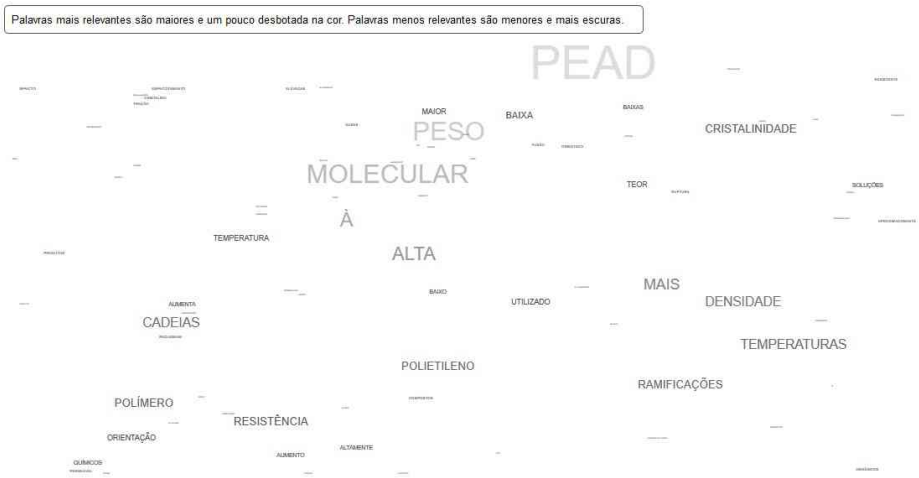

Figura 5 - Word Cloud (Nuvem de Palavras) para um estudante em uma atividade.

O *TreeMap* (Figura 7), apresenta a análise de todas as atividades de um estudante. As caixas da *TreeMap* representam os polímeros sob análise. A cor das caixas indica a porcentagem de similaridade que cada polímero possui em relação ao trecho selecionado. No canto superior direito da tela há uma legenda que associa a cor da caixa com o tipo do polímero. Quanto maior a pontuação de uma palavra, maior a área ocupada (e vice-versa). Posicionando o cursor sobre a área da palavra, pode ser visualizada a palavra por completo acompanhada da sua pontuação. O professor avaliou que esta visualização é mais eficiente que o gráfico das bolhas (Figura 4), pois as bolhas geram uma distorção circular na visualização. Segundo o professor, isto não ocorre quando se utiliza quadriláteros.

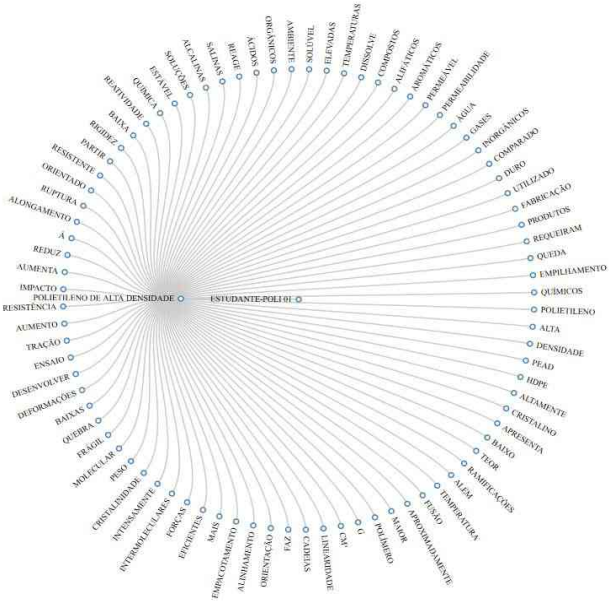

Figura 6 - Rotating Cluster (Agrupamento Rotativo) de um estudante em uma atividade.

A Figura 8 apresenta o Bar Chart que permite a visualização da avaliação quantitativa com relação a quantidade de palavras relevantes utilizadas pelos estudantes. Cada barra se refere a um estudante da turma. Posicionando-se o cursor sobre a barra, tem-se a informação da frequência, que é justamente a quantidade de palavras relevantes para a atividade analisada. O objetivo principal desta visualização é comparar relativamente o quanto cada estudante se aproximou do texto original. Neste sentido, os estudantes 1, 6 e 7 apresentaram melhores frequências com relação ao número de palavras significativas dos seus textos.

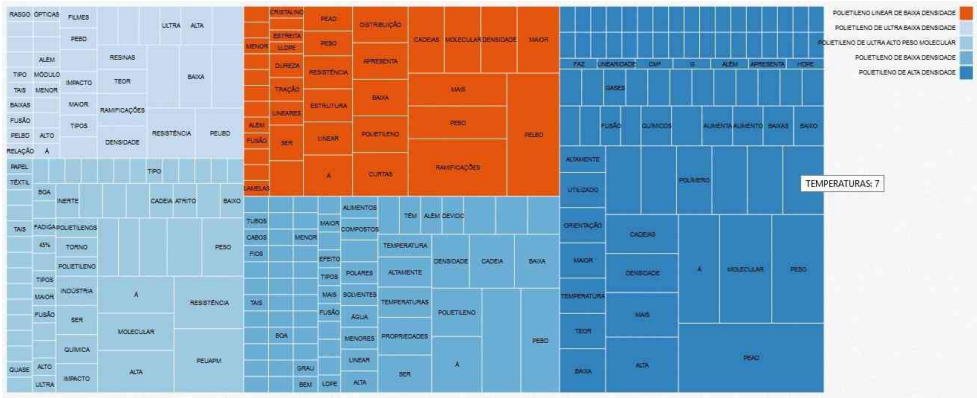

Figura 7 - Treemap (Mapa Árvore) de um estudante em todas as atividades.

A Figura 9 apresenta o LSA ScatterPlotBar, relacionando os resultados obtidos pelo algoritmo LSA com a produção textual dos alunos. Este gráfico mostra graficamente os valores produzidos pela aplicação do LSA às produções dos alunos. É possível observar a distribuição das produções textuais no espaço semântico, onde a diferença na qualidade de produção entre os estudantes de maior e menor nota são evidenciadas.

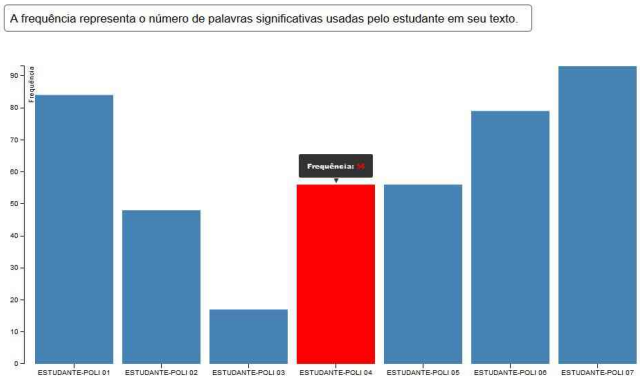

Figura 8 - Bar Chart (Gráfico de Barras) para uma turma em uma atividade.

Na análise do professor, os resultados obtidos foram satisfatórios, pois o sistema é capaz de avaliar a produção textual dos alunos e qualificá-los com referência ao texto utilizado no treinamento.

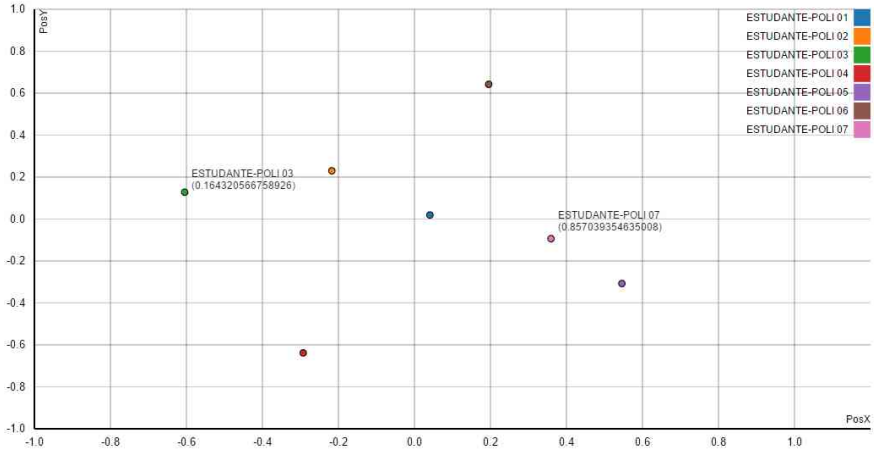

Figura 9 - LSA ScatterPlotBar para uma turma

A nota que o algoritmo atribui (coordenada x,y) não deve ser utilizada como nota absoluta, uma vez que pode haver notas negativas no resultado final. No entanto, podem

ser utilizadas para classificar as produções textuais entre si.

5. Conclusões Preliminares

O aprendizado de máquina oferece técnicas úteis na extração do conhecimento em textos, identificando palavras, expressões e termos relevantes em uma produção textual. Porém, para facilitar a análise de um texto é importante a apresentação visual destes dados. Foi então desenvolvido o software APT para a análise e visualização de produções textuais para uso educacional em qualquer área do conhecimento.

O primeiro experimento no uso do software foi realizado na disciplina de Química Orgânica. A avaliação preliminar feita pelo professor responsável pelo experimento identificou os seguintes pontos positivos no uso da ferramenta APT:

(a) reconhecimento do vocabulário empregado pelos estudantes
(b) ponderação do nível de coerência do texto do aluno em relação ao corpus do professor
(c) percepção do desenvolvimento textual global da turma
(d) comparação do nível de redação entre alunos da mesma turma
(e) detecção das relações mais fortes entre conceitos do texto original
(f) pontos de maior compreensão e menor compreensão do texto original

Como se trata de um projeto em andamento, novos algoritmos de aprendizado de máquina para análise das produções textuais se fazem necessários para que se produza resultados que auxiliem e contribuam ainda mais com a tarefa de leitura e avaliação textual. Para enriquecer a visualização dos dados minerados também são necessários novos produtos que proponham melhorar a interatividade do sistema. Para finalizar, conclui-se que os métodos automáticos podem operar sobre as produções textuais e produzir análises que forneçam ao professor visões sobre a evolução dos estudantes e da turma como um todo.

6. Referências Bibliográficas

ARANHA, C. N. Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional. Rio de Janeiro, 2007.

BERRY, M. W.; DRMAC, Z.; JESSUP, E. R. Matrices, Vector Spaces, andInformation Retrieval. SIAM Rev. Philadelphia, PA. 1999.

BOSTOCK, M. D3 Data-Driven Documents. D3 Data-Driven Documents, 2013. Disponível em: <http://d3js.org/>. Acesso em: 22 junho 2015.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. Information Visualization. Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers. San Francisco, 1999.

COUTINHO, F. M. B.; MELLO, I. L.; MARIA, L. C. D. S. Polietileno: Principais Tipos, Propriedades e Aplicações. Polímeros: Ciência e Tecnologia, Rio de Janeiro, v. 13, n. 1, 2003.

DEERWESTER, S. E. A. Indexing by latent semantic analysis. Journal of the American Society for Information Science. [S.l.]. 1990.

FELDAMAN, R., SANGER, J. Text Mining Handbook. Cambridge, Inglaterra: Cambridge University Press, 2006.

FREITAS, C. M. D. S.; CHUBACHI, O. M.; LUZZARDI, P. R. G.; CAVA R. A. Introdução à Visualização de Informações. RITA – Revista de Informática Teórica e Aplicada, Instituto de Informática, v. 8, p. 143-158, 2001.

IBM Watson. Disponível em: <http://www.ibm.com/watson/>. Acesso em: 22 setembro 2016.

JUDELMAN, G. B. Knowledge Visualization: Problems and Principles for Mapping the Knowledge Space. Germany, 2004.

KEIM, D. A. Information visualization and visual data mining. IEEE Trans. on Visualization and Comp. Graphics, v. 8, n. 1, p. 1–8, 2002.

MICROSOFT. Cognitive Services. Disponível em: <https://www.microsoft.com/cognitive-services/en-us/apis>. Acesso em: 10 setembro 2016.

MANNING, C. D., RAGHAVAN P., SCHUTZE, H., An Introduction to Information Retrieval. Cambridge University Press, 2008

TAN, A.-H. Text mining: The state of the art and the challenges. In: PAKDD 1999 WORKSHOP ON KNOWLEDGE DISOCOVERY FROM ADVANCED DATABASES. Beijing, 1999.

WEBBER, C.G., CINI, G., LIMA, M.F.W.P. Facilitando a Análise se Dados Educacionais através se Ferramentas se Visualização. Revista Novas Tecnologias na Educação, RENOTE, v. 11, n. 3, 2013.