2. Почему мы используем F-меру, почему, например, нельзя просто взять среднее от Precision и Recall?

Precision и Recall дают довольно исчерпывающую характеристику классификатора, причем «с разных углов». Обычно при построении подобного рода систем приходится все время балансировать между двумя этими метриками. Если вы пытаетесь повысить Recall, делая классификатор более «оптимистичным», это приводит к падению Precision из-за увеличения числа ложно-положительных ответов. Если же вы подкручиваете свой классификатор, делая его более «пессимистичным», например, строже фильтруя результаты, то при росте Precision это вызовет одновременное падение Recall из-за отбраковки какого-то числа правильных ответов. Поэтому удобно для характеристики классификатора использовать одну величину, так называемую метрику F1.

