

ActivitiesVisual Studio Code

FileEditSelectionViewGoRunTerminalHelp

EXPLORERPADDLE-PLAYGROUND> any> markdowns6MLH-display-case-sales-sheet-2.mdimage.md> pdfs6MLH-display-case-sales-sheet-2.pdfimage.pdfSPOILER\_image\_cropped184.pdf> scripts> \_\_pycache\_\_start.py> src> \_\_pycache\_\_> services> \_\_pycache\_\_any\_to\_pdfs.pyimage\_to\_text.pypdf\_to\_image.pypdf\_to\_markdowns.py> main.py> textsimage copy.txtimage.txtSPOILER\_image\_cropped184.txt.gitignorepoetry.lockpyproject.tomlREADME.md

pdf\_to\_markdowns.py M

pymupdf\_rag.py X

pdf\_to\_image.py M

6MLH-display-case-sales-sheet-2.pdf U

any\_to\_pdfs.py U

6MLH-display-case-sales-sheet-2.md

n

home > vikiw > .cache > pypoetry > virtualenvs > src--rG\_ktix-py3.11 > lib > python3.11 > site-packages > pdf4llm > helpers > pymupdf\_rag.py > to\_markdown

47def to\_markdown(doc: fitz.Document, pages: list = None) -> str:

287text\_rects.append(("text", page.rect, 0))

288else:

289rtype, r, idx = text\_rects[-1]

290if rtype == "table":

291tr = page.rect

292tr.y0 = r.y1

293if not tr.is\_empty:

294text\_rects.append(("text", tr, 0))

295

296# we have all rectangles and can start outputting their contents

297for rtype, r, idx in text\_rects:

298if rtype == "text": # a text rectangle

299md\_string += write\_text(page, r, hdr\_prefix) # write MD content

300md\_string += "\n"

301else: # a table rect

302md\_string += tabs[idx].to\_markdown(clean=False)

303

304md\_string += "\n----\n\n"

305

306return md\_string

307

308

309if \_\_name\_\_ == "\_\_main\_\_":

310import os

311import sys

312import time

313

314try:

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

COMMENTS

bash

return loop.run\_until\_complete(task)

~~~~~

File "/home/vikiw/.cache/pypoetry/virtualenvs/src--rG\_ktix-py3.11/lib/python3.11/site-packages/nest\_asyncio.py", line 98, in run\_until\_complete

return f.result()

~~~~~

File "/usr/lib/python3.11/asyncio/futures.py", line 203, in result

raise self.\_exception.with\_traceback(self.\_exception\_tb)

File "/usr/lib/python3.11/asyncio/tasks.py", line 277, in \_\_step

result = coro.send(None)

~~~~~

File "/home/vikiw/Desktop/paddle-playground/src/services/pdf\_to\_markdowns.py", line 17, in arun

md\_text = pdf4llm.to\_markdown(doc=file\_path, pages=[0])

~~~~~

File "/home/vikiw/.cache/pypoetry/virtualenvs/src--rG\_ktix-py3.11/lib/python3.11/site-packages/pdf4llm/helpers/pymupdf\_rag.py", line 238, in to\_markdown

hdr\_prefix = IdentifyHeaders(doc, pages=pages)

~~~~~

File "/home/vikiw/.cache/pypoetry/virtualenvs/src--rG\_ktix-py3.11/lib/python3.11/site-packages/pdf4llm/helpers/pymupdf\_rag.py", line 85, in \_\_init\_\_

body\_limit = sorted()

~~~~~

IndexError: list index out of range

src-py3.11vikiw@vikiw-X456UR:~/Desktop/paddle-playground\$

Ln 47, Col 5Spaces: 4UTF-8CRLFPython 3.11.9 ('src--rG\_ktix-py3.11': Poetry)Codeium: {...}